In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institute shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

7/25/68

NONLINEAR PROGRAMMING: NONDIFFERENTIABLE FUNCTIONS

A THESIS

Presented to

The Faculty of the Division of Graduate

Studies and Research

by

Mokhtar Sadek Bazaraa

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

in the School of Industrial and Systems Engineering

Georgia Institute of Technology

December, 1969

NONLINEAR PROGRAMMING: NONDIFFERENTIABLE FUNCTIONS

Approved:

Chairman

Date approved by Chairman: December 13, 1969

ACKNOWLEDGMENTS

I gratefully acknowledge the assistance, counsel, and friendship of my advisor, Dr. C. M. Shetty. Without his help, valuable suggestions, and deep insight, this research would not have been possible, and I would like to express my appreciation to him.

Special thanks are due to Dr. J. J. Goode. Dr. Goode has helped the author throughout the study by his valuable suggestions. He has been always approachable, and his mathematical ability and insight have been extremely valuable.

Thanks are due to Dr. D. E. Fyffe for his helpful suggestions and criticism. I would like also to thank Dr. E. R. Immel and Dr. V. E. Unger for their help in reading the dissertation.

I also would like to express my deep thanks to Dr. R. N. Lehrer, Director of the School of Industrial and Systems Engineering, Georgia Tech, for encouraging me throughout and giving me every possible assistance.

Special thanks are due to Dr. W. W. Hines, Associate Director, Graduate Programs, School of Industrial and Systems Engineering, Georgia Tech. Dr. Hines has guided me all throughout my study and his inspiration and assistance I greatly appreciate.

Finally I would like to express my thanks to Mrs. Betty Sims for an excellent typing job.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

A nonlinear programming problem is usually expressed in the form: Find an $x = (x_1, x_2, \ldots, x_n)$ so as to minimize the objective (criterion) function $f_o(x)$, subject to the constraints $f_i(x) \leq 0$, $i = 1,2,\ldots,m$. Such problems frequently arise in practice and a good discussion on the applications of nonlinear programming is found in [6], [31], and [58]. In view of the importance of the problem, considerable research effort has been directed in recent years to both theoretical and computational aspects of nonlinear programming. This study is mainly concerned with the former, and discusses in depth optimality criteria and duality in nonlinear programming.

In a nonlinear programming problem, when the functions involved are differentiable, the frequently used necessary conditions for optimality are the well known Kuhn-Tucker [30] and Fritz John conditions [27]. These conditions are also sufficient, under certain assumptions which are mainly associated with convexity. One of the objective of this study is to investigate the conditions for optimality when differentiability, continuity, and convexity assumptions are relaxed. The second major objective of the study is to unify various duality formulations in nonlinear programming so that they are all subsumed under one formulation. Some new results in this connection along with economic and geometric interpretation of duality are also investigated.

We will now discuss in slightly more detail the scope of the study and the organization of the presentation of the results. We then present the notation adopted throughout the study. The last section in this chapter discusses some situations when the results of this study would be relevant.

1. Scope of the Study

Most studies in nonlinear programming assume that the functions involved are differentiable. Throughout this study such an assumption is relaxed, which implies that the existence of the gradient vector is not assured. In Chapter II we introduce the known notion of outernormals to sets, and using this concept we define subgradients to functions. When the functions under consideration are differentiable, subgradients are identical to the gradient vectors. Thus the introduction of subgradients emphasizes the similarity between differentiable and nondifferentiable[†] functions.

At this stage we also introduce the concept of supportable functions which play the role of convexity in some of the important theorems in Chapter II. We then discuss different types of convexity and investigate their relationship with subgradients. This is followed by definitions of minimal cones and dual cones of sets, and their relationship with outernormals.

Considering the sets above and below the functions involved in a

---

[†]Throughout the study, by a "nondifferentiable function" we mean a function which is not necessarily differentiable everywhere.

nonlinear programming problem, and their associated cones lead to some interesting results. It permits us to partially relax the convexity assumption, which is replaced by supportability of the functions under consideration. It also permits us to relax the continuity assumption in some cases.

The important conditions for optimality developed in Chapter II, and the major corresponding assumptions may be summarized as follows.

(i) Necessary conditions when the objective function and the constraint functions are continuous, nondifferentiable, and either locally supportable from below or above. These conditions are similar to the Fritz John necessary conditions.

(ii) Necessary and sufficient conditions when the objective function and the constraint functions are continuous, but not necessarily differentiable, and are all locally supportable from below. It is also assumed that the constraint set has an interior point. These conditions are similar to the Kuhn-Tucker conditions.

(iii) Necessary and sufficient condition for optimality of an interior point when the constraint function are continuous but not necessarily differentiable.

(iv) A separate set of sufficient conditions when the objective function is locally supportable from below. No restrictions regarding the constraint functions are imposed.

We now turn to Chapter III, where duality in nonlinear programming is considered. First, it will be shown that existing duality formulations are subsumed under the Minmax formulation. In particular, Fenchel's formulation of duality via conjugate functions, Falk's formulation of duality via the lagrangian multiplier vector and the symmetric duality formulation of Dantzig, Eisenberg and Cottle (which together dominate other duality formulations) are derived from the Minmax formulation. In the process we obtain extensions of some important duality theorems by relaxing differentiability and partially relaxing continuity and convexity. We also give economic and geometric interpretations of duality. The final chapter of this study summarizes the results and indicates some areas for further research.

We now present the conventions and notation used in this study. The n dimensional Euclidean space is denoted by E^n. The non-negative orthant of E^n is denoted by E^n_+. If x belongs to E^n ($x \varepsilon E^n$), then the components of x are denoted by attaching subscripts to x. Superscripts are used to denote different vectors. The ordering relation "$\leq$" in E^n is defined such that $x \leq y$ if and only if (iff) $x_i \leq y_i$ for i = 1,2,...,n. Similarly, $x < y$ iff $x_i < y_i$ for i = 1,2,...,n. The number zero and the zero vector are denoted by the same character o.

Subsets of E^n are denoted by upper case Latin letters, and elements of these subsets are denoted by lower case Latin letters. Scalars are usually denoted by lower case Greek letters. Lower case Greek letters may also be used to denote functions. The difference is clear from the context.

Sets are defined either by explicitly mentioning all elements of the set or by stating the properties which define the set, e.g. A = {x: x satisfies Q} is used to define the set A, consisting of all

x satisfying property Q. If the set B is a subset of A (possibly equal to A) then we denote it by $B \subset A$. We reserve the letter C for cones and denote dual cones by C^*.

The closure of a set A is denoted by $\bar{A}$. The boundary of a set A is denoted by ∂A and the convex hull of a set A is denoted by $[A]$. The interior of a set A is denoted by $\mathrm{int}(A)$ and the relative interior of A is denoted by $r(A)$. Outernormals and local outernormals to sets are abbreviated to *o.n.* and *l.o.n.* and are usually denoted by p.

The set $\{x: x \varepsilon E^1, a \leq x \leq b\}$ is denoted by $[a,b]$, and the set $\{x: x \varepsilon E^1, a < x < b\}$ is denoted by (a,b). Similarly, the set $\{x: x \varepsilon E^1, a < x \leq b\}$ is denoted by $(a,b]$.

The cartesian product of two sets E and F, i.e. the set $\{(x,y): x \varepsilon E, y \varepsilon F\}$ is denoted by $E x F$.

A function $f: X \rightarrow Y$ is a single valued mapping, with domain $X \subset E^n$ and range contained in $Y \subset E^m$. Functions are denoted by either lower case Latin or lower case Greek letters. If $Y \subset E^1$, we denote the gradient vector of the function f at $\bar{x} \varepsilon X$ by $\nabla f(\bar{x})$. Similarly a subgradient of f at $\bar{x}$ is denoted by $\theta(\bar{x})$.

The inner product of two vectors x and $y \varepsilon E^n$ is denoted by $<x,y>$ and is given by $\sum_{i=1}^{n} x_i y_i$. The norm of a vector $x \varepsilon E^n$ is denoted by $\|x\|$ and is equal to $\sqrt{<x,x>}$.

A δ-neighborhood of a point $\bar{x} \varepsilon E^n$ is denoted by $N(\bar{x},\delta)$, and given by the set $\{x: x \varepsilon E^n, \|x-\bar{x}\| < \delta\}$. Usually the radius δ is of no importance, and the neighborhood is simply denoted by $N(\bar{x})$ or N.

Theorems, remarks, definitions, and equations are numbered consecutively within each chapter. Whenever a reference is made, unless otherwise stated, we mean a theorem, remark, definition, or an equation in the same chapter.

2. Relevance of the Study

This study is mainly devoted to optimality criteria and duality in nonlinear programming. Throughout the study we assume that the functions involved are not necessarily differentiable. Furthermore, in developing the optimality criteria we assume that the objective function is supportable from below (or above) but not necessarily convex. We also partially relax the convexity assumption of the constraints to supportable functions in some instances, and completely relax the assumption in other cases. The assumption of continuity of the functions involved is also relaxed in some cases.

In many practical problems one frequently encounters nondifferentiable functions which may or may not be continuous and which may or may not be convex. Even though the functions involved in a realistic problem may be differentiable in the region of interest, it is unlikely that differentiability and continuity properties are satisfied everywhere. In yet other instances, nondifferentiability, noncontinuity, and nonconvexity may arise naturally in the problem. Some of these situations are stated below to illustrate the point.

Nondifferentiability of a function f_o arises in a natural way if the function is of the form, $f_o = \max(f_o^1, f_o^2, \ldots, f_o^k)$, where the original functions $f_o^i (i=1,2,\ldots,k)$ may or may not be differentiable.

This, in fact, renders all piecewise convex (and piecewise concave) programs to be nondifferentiable. See, for example, Zangwill [57].

Likewise, in Minmax theory nondifferentiable functions arise in a natural way. Suppose that the function under consideration ϕ: $E \times F \to E^1$, where $E \subset E^n$ and $F \subset E^m$ is differentiable. The two functions α and β defined below are not necessarily differentiable:

$$\alpha(x) = \sup_y \{\phi(x,y)\colon y \in F\} \qquad \text{for all } x \in E$$

$$\beta(y) = \inf_x \{\phi(x,y)\colon x \in E\} \qquad \text{for all } y \in F$$

The importance of such functions, e.g. in military applications, is discussed by Danskin [12]. We will discuss in Chapter III in more detail the above formulation which is the basis of our dual programs.

Another example where nondifferentiability arises naturally, along with nonconvexity and semi-continuity, is when a "fixed charge" or a "price break" is applied based on the level of the activity vector x. See, for example, [10] and [25].

We now turn to the relevance of the results presented in this study. Optimality criteria and duality theorems have been used in various forms to develop practical solution procedures and are discussed below. It is our hope that the results presented in Chapters II and III of the study will facilitate development and extension of algorithms to problems for which there are no suitable solution procedures at present.

We discuss below some solution procedures which are based on the optimality criteria and/or duality. We also discuss some algorithms which are less dependent upon optimality criteria and/or duality, and which use them essentially as a stopping rule or as a subroutine of the solution procedure.

A natural class of solution procedures which is based on the Kuhn-Tucker conditions is quadratic programming. The reason is obvious; namely, the conditions for optimality are linear in the decision vector x, and hence the simplex method of linear programming (with slight modifications) can be used to find a solution that satisfies the Kuhn-Tucker conditions. This further implies that the solution is an optimal solution of the quadratic program under consideration.

Different algorithms in quadratic programming which are based on the Kuhn-Tucker conditions exist, such as the algorithms of Wolfe [55], Hildreth [24], and Barankin and Dorfman [4]. The following is a brief discussion of Wolfe's procedure which appears to be the most efficient and well known algorithm in quadratic programming, among those that are based on the Kuhn-Tucker conditions.

The problem under consideration is to minimize a quadratic objective function, subject to linear constraints, i.e. $\text{minimize}_x\{\langle q,x\rangle + \langle x,Mx\rangle : x \geq o,\ \langle a_i,x\rangle = b_i,\ i = 1,2,\ldots,m\}$ where q is an n dimensional vector, M is an nxn positive semidefinite matrix, a_i $(i=1,2,\ldots,m)$ are n dimensional vectors, and b_i $(i=1,2,\ldots,m)$ are scalars. This assures that the objective function is convex and that the feasible domain is a convex set. The Kuhn-Tucker conditions (discussed in Chapter II) for this problem are,

$$\langle a_i, x\rangle = b_i \qquad i=1,2,\ldots,m.$$

$$2Mx + \sum_{i=1}^{m} u_i a_i - v = -q$$

$$x \geq o,\ v \geq o,\ \langle x, v\rangle = o.$$

Hence it suffices to find a solution of the above system of equalities and inequalities. In general terms, Wolfe's algorithm consists of constructing an extended system of equations by introducing a set of artificial variables to obtain a basic solution. These artificial variables are then made to vanish in phase I of the simplex method of linear programming. Through an additional rule for the transition from one basic solution to the next, we are assured that the condition $\langle x,v\rangle = o$ is satisfied. For further details, one may refer to [55].

In other algorithms, the theorems that validate the procedure may be based on the Kuhn-Tucker conditions, even though the latter may not appear explicitly in the steps of the solution procedure. An example of this class of algorithms is the Theil-Van de Panne algorithm for quadratic programs. See, for example, [50] and [31]. This algorithm is validated by the following theorems, which in turn could be based on the Kuhn-Tucker conditions. The problem under consideration, denoted by P, is of the form,

$$\text{minimize}_x\{\langle q,x\rangle + \langle x,Mx\rangle : \langle a_i,x\rangle \le b_i, \quad i=1,2,\ldots,m\}$$

where q is an n dimensional vector, M is an nxn positive definite matrix, a_i $(i=1,2,\ldots,m)$ are n dimensional vectors, and b_i $(i=1,2,\ldots,m)$ are scalars. In the theorems and discussion below, $x(S)$ denotes a solution of the problem $P(S)$,

$$\text{minimize}_x\{\langle q,x\rangle + \langle x,Mx\rangle : \langle a_i,x\rangle = b_i, \quad \text{for all } i\varepsilon S\}$$

and $V(x(S))$ denotes the set $\{i : \langle a_i,x(S)\rangle > b_i\}$.

Theorem 1

Let $\bar{x}$ be an optimal solution of problem P, and let $\bar{S} = \{i : \langle a_i,\bar{x}\rangle = b_i\}$. Then $\bar{x} = x(\bar{S})$.

Theorem 2

If $\bar{x} = x(\bar{S})$ solves problem P, and if $\bar{S} \neq \phi$, then for all $S\subset\bar{S}$ and $S \neq \bar{S}$, $h\varepsilon V(x(S))$ for some $h\varepsilon\bar{S}-S$.

Theorem 3

If $V(x(S)) = \phi$ and $h\varepsilon V(x(S-\{h\}))$ for all $h\varepsilon S$, then $x(S)$ is the optimal solution of problem P.

The following gives a compact description of the algorithm. It should be noted that at each stage j of the algorithm, a number of equality constrained problems is solved, where the number of constraints in each problem is precisely j. The algorithm is based on Theorems 1, 2, and 3 above.

(i) Let j=o and solve the problem, $\text{minimize}_x\{\langle q,x\rangle + \langle x,Mx\rangle: x\varepsilon E^n\}$. Let x^o be an optimal solution. If x^o is a feasible solution of problem P, then stop, otherwise let $\mathcal{S}_o = \{S_o: S_o = \{i\}, \langle a_i,x^o\rangle > o\}$.

(ii) Let $j = j+1$.

(iii) Choose $S_{j-1} \varepsilon \mathcal{S}_{j-1}$, solve the problem $P(S_{j-1})$, and let $x(S_{j-1})$ be an optimal solution. If $V(x(S_{j-1})) = \phi$, then check optimality via Theorem 3. If $x(S_{j-1})$ solves problem P, then stop, otherwise $P(S_{j-1})$ is dropped from any further consideration. If on the other hand $V(x(S_{j-1})) \neq \phi$, then form the collection of sets of indices $\hat{\mathcal{S}}_j(S_{j-1}) = \{S_{j-1}\cup\{i\}: i\varepsilon V(x(S_{j-1}))\}$.

(iv) Repeat step (iii) above for all $S_{j-1} \varepsilon \mathcal{S}_{j-1}$ and then form the new collection $\mathcal{S}_j = \{S_j: S_j \varepsilon \hat{\mathcal{S}}_j(S_{j-1}), S_{j-1} \varepsilon \mathcal{S}_{j-1}\}$. Then go to step (ii).

In Appendix A we give a generalization of Theorems 1 and 2 above when the objective function is strictly convex and the constraint functions are convex. We also present a brief outline of an algorithm that solves such problems. Basically the algorithm is a "Branch and Bound" procedure which makes use of Theorem 2 of the Appendix.

We now present some instances where the Kuhn-Tucker conditions are used indirectly either as a stopping rule or as a subroutine in various gradient methods for solving nonlinear programs. The problem under consideration is of the form,

$$\text{minimize}_x\{f_o(x): f_i(x) \leq o, \quad i=1,2,\ldots,m\}.$$

Gradient methods can be summarized in the following steps.

(i) Suppose that the feasible solutions $x^1, x^2, \ldots, x^k$ have already been calculated.

(ii) Starting from x^k, determine a direction s^k (related to the gradient) with the property that there exists $\bar{\lambda} > o$ such that for all $\lambda \varepsilon (o, \bar{\lambda}]$, $f_o(x_\lambda^k) < f_o(x^k)$, where $x_\lambda^k = x^k + \lambda s^k$.

(iii) If x_λ^k is feasible for all $\lambda \varepsilon (o, \bar{\lambda}]$, then determine the step-size λ_k, e.g. by solving the one dimensional minimization problem in λ, $\text{minimize}_\lambda \{f_o(x_\lambda^k) : f_i(x_\lambda^k) \leq o,\ i=1,2,\ldots,m\}$.

The solution of this problem is denoted by x^{k+1}. If, on the other hand, x_λ^k is not feasible for any $\lambda \varepsilon (o, \bar{\lambda}]$, then choose $x_{\lambda_o}^k$ corresponding to $\lambda_o \varepsilon (o, \bar{\lambda}]$. Starting from $x_{\lambda_o}^k$, a move is determined leading to x^{k+1} in the feasible region, such that $f_o(x^{k+1}) < f_o(x_{\lambda_o}^k)$. In either case the new feasible solution x^{k+1} is an improved solution.

(iv) Step (ii) is repeated until an optimal solution is obtained.

Different methods of direction finding and different methods of determining suitable stepsize λ may lead to different algorithms. To avoid "zigzagging," a common feature in most gradient methods, and to guarantee or speed up convergence, additional requirements can be added to the process of direction finding.

One of the rules for determining a direction s given by Zoutendijk [59] makes use of the Kuhn-Tucker conditions. Suppose that a direction s^k is to be found starting from the feasible solution x^k, where $f_i(x^k) = o$ for all $i \varepsilon I \subset \{1,2,\ldots,m\}$. The problem can be stated in the form,

$$\text{minimize}_s\{\langle\nabla f_o(x^k),s\rangle: \langle\nabla f_i(x^k),s\rangle \le o \text{ for all } i\epsilon I, \langle s,s\rangle \le 1\}.$$

The constraints $\langle\nabla f_i(x^k),s\rangle \le o$ for all $i\epsilon I$ would guarantee that a movement in the direction s starting from x^k may lead to a feasible solution for a sufficiently small stepsize λ, while the constraint $\langle s,s\rangle \le 1$ is convenient since we are mainly interested in the direction. If $\nabla f_o(x^k) \ne o$, then the minimization of $\langle\nabla f_o(x^k),s\rangle$ would guarantee an improvement in the value of the objective function if we move in direction s^k iff $\langle\nabla f_o(x^k),s^k\rangle < o$, where s^k results from the minimization process mentioned above. This leads to an obvious stopping rule; namely, the process is terminated if the minimization above generates a direction s^k such that $\langle\nabla f_o(x^k),s^k\rangle = o$.

The Kuhn-Tucker conditions (discussed in Chapter II) for this problem can be used to find the direction s^k. These conditions may be written in the form,

$$-\nabla f_o(x^k) = A_k' u + u_o s^k$$

$$A_k s^k + y = o$$

$$\langle u,y\rangle = o, \quad u \ge o, \quad u_o \ge o, \quad y \ge o.$$

where A_k is an rxn matrix formed by the gradient vectors $\nabla f_i(x^k)$ where $i\epsilon I = \{i_1,i_2,\ldots,i_r\}$, A_k' is the transpose of A_k, u and y are r dimensional vectors, and u_o is a scalar. The above conditions may be rewritten in the form,

$$A_k \nabla f_o(x^k) = -A_k A_k' u + v$$

$$u \geq o, \; v \geq o, \; \langle u, v \rangle = o$$

where $v = u_o y = -u_o A_k s^k$.

Zoutendijk proposed a procedure using the dual simplex method to solve the above system. Another procedure to find a direction s is to use the Kuhn-Tucker conditions to formulate a simple equivalent quadratic program which may be solved by Wolfe's procedure discussed earlier. For further details one may refer to [31].

Again in Rosen's gradient projection method [46] for nonlinear programming problems with linear constraints, the Kuhn-Tucker conditions are used as a stopping rule. Suppose that the problem under consideration is of the form,

$$\text{minimize}_x \{f_o(x): \langle a_i, x \rangle \leq b_i, \quad i=1,2,\ldots,m\}$$

and suppose further that x^k is a given feasible solution of this problem. If x^k is an interior point of the feasible region, then the direction of movement s^k is taken to be $-\nabla f_o(x^k)$. On the other hand, if $\langle a_i, x^k \rangle = b_i$ for some $i \in \{1,2,\ldots,m\}$, then s^k is given by the projection of $-\nabla f_o(x^k)$ on the affine manifold L formed by the intersection of the hyperplanes $\langle a_i, x \rangle = b_i$ for $i \in I = \{i_1, i_2, \ldots, i_r\} = \{i: \langle a_i, x^k \rangle = b_i\}$. The vectors a_i for all $i \in I$ are orthogonal to the affine manifold L and they span an orthogonal affine manifold $\tilde{L}$.

It can be shown that the projection matrices into L_k and $\tilde{L}_k$ are given by $P_k = I_n - A_k(A_k'A_k)^{-1}A_k'$ and $\tilde{P}_k = A_k(A_k'A_k)^{-1}A_k'$, where I_n is the nxn identity matrix, and A_k' is an rxn matrix formed by $a_{i_1}, a_{i_2}, \ldots, a_{i_r}$ as its rows. (The assumption that $a_{i_1}, a_{i_2}, \ldots, a_{i_r}$ are independent is implicit. If not then choose the maximum number of independent vectors among $a_{i_1}, \ldots a_{i_r}$ to form A_k'). For a proof of the above statements one may refer to [31].

The stopping rule given by Rosen is to terminate the process when a feasible solution x^{k_o} is found such that, $P_{k_o} \nabla f_o(x^{k_o}) = o$, and $(A_{k_o}' A_{k_o})^{-1} A_{k_o}' \nabla f_o(x^{k_o}) \leq o$. However, these conditions are precisely the Kuhn-Tucker conditions. This is immediate by noting that since $P_{k_o} \nabla f_o(x^{k_o}) = o$, then $\nabla f_o(x^{k_o})$ is orthogonal to L_{k_o} and hence can be represented as a linear combination of $a_{i_1}, a_{i_2}, \ldots,$ and a_{i_r}. In other words, $-\nabla f_o(x^{k_o}) = A_{k_o} u$ for some r dimensional vector u. This further implies that the above conditions may be rewritten as,

$$-\nabla f_o(x^{k_o}) = \sum_{i \in I} u_i a_i, \quad u_i \geq o \quad \text{for all } i \in I.$$

This is the usual form of the Kuhn-Tucker conditions. Likewise, duality may be fruitfully utilized in developing solution procedures. We discuss some solution procedures which are based on duality. We first consider the generalized lagrangian multiplier technique due to Everett [16]. Assume that the problem under consideration is of the form,

$$\text{minimize}_x\{f_o(x): f_i(x) \le o, \quad i=1,2,\ldots,m\}.$$

Everett asserts that if $x^o \in E^n$ solves the unconstrained problem, $\text{minimize}_x\{f_o(x) + \langle u,f(x)\rangle: x \in E^n\}$, where u is a nonnegative m dimensional vector, and $f = (f_1,f_2,\ldots,f_m)$, then x^o solves the problem,

$$\text{minimize}_x\{f_o(x): f_i(x) \le f_i(x^o), \quad i=1,2,\ldots,m\}.$$

Therefore if we can find a lagrangian vector $u^o \ge o$ such that $f_i(x^o) = o$ for $i=1,2,\ldots,m$, then x^o solves the original problem. This fact is used to change the problem into a sequence of unconstrained minimization problems, which may be solved by one of the existing powerful solution procedures. See for example [21], [23], and [53]. The above procedure can be viewed as a dual procedure, as we indicate in Chapter III.

Falk [17] has introduced a duality formulation by the introduction of a lagrangian multiplier vector, which is similar to Everett's. The results of Falk may be utilized to solve a nonlinear programming problem. Assume that we want to minimize a function subject to both linear and nonlinear inequality constraints. The problem is thus of the form,

$$\text{minimize}_x\{f_o(x): f_i(x) \le o, \quad i=1,2,\ldots,m; \quad Ax \le b\}$$

where A is an rxn matrix and b is an r dimensional vector.

Introducing the nonlinear constraints into the objective function, the function β: $E^m \to E^1$ can be defined as $\beta(u) = \inf_x\{f_o(x) + \langle u,f(x)\rangle$: $Ax \le b\}$ for $u \ge 0$. The problem, $\text{maximize}_u\{\beta(u)\text{: } u \ge 0\}$ is considered. Under some restrictions of the functions $f_o,f_1,\ldots,f_m$ it can be shown that the solutions of the original problem and the above problem (dual problem) are equivalent. It may be easier to solve the dual program, which consists of maximizing a concave function, subject to nonnegativity constraints, than to solve the original problem.

Another example where duality is used as a solution procedure is the penalty function technique [21] and [58]. This technique can be viewed as a primal-dual approach. See, for example, [20], [45], and [58]. To clarify this further, consider the function ϕ: $ExF \to E^1$, where $E \subset E^n$ and $F \subset E^r$, and let $\alpha(x) = \sup_y\{\phi(x,y)\text{: } y \in F\}$ for all $x \in E$, and $\beta(y) = \inf_x\{\phi(x,y)\text{: } x \in E\}$ for all $y \in F$. Let problems P and D be defined by:

$$\text{P: minimize}_x\{\alpha(x)\text{: } x \in E\}$$

$$\text{D: maximize}_y\{\beta(y)\text{: } y \in F\}$$

We now specialize the set E to be E^n and let the function ϕ be of the form,

$$\phi(x,y) = f_o(x) + h(y,f_1(x),f_2(x),\ldots,f_m(x))$$

$$\text{for } x \in E^n,\ y \in F \subset E^r,$$

where F contains E_+^r. We further restrict the function h to satisfy the following properties.

(i) If $x \epsilon E^n$ and $f_i(x) \leq 0$, $i=1,2,\ldots,m$, then $h(.,f_1(x),f_2(x),\ldots,f_m(x)) = 0$ for all $y \epsilon F$.

(ii) If $x \epsilon E^n$ and $f_i(x) > 0$ for some $i \epsilon \{1,2,\ldots,m\}$, and if $\{y^k\}$ is an increasing sequence in F, then $\{h(y^k,f_1(x),\ldots,f_m(x))\}$ is an increasing sequence in E^1. Moreover, if $\lim_{k\to\infty} y^k = \infty$†, then $\lim_{k\to\infty} h(y^k,f_1(x),\ldots,f_n(x)) = \infty$.

Under such a specialization of the function ϕ, we get:

$$\alpha(x) = \sup_y\{f_o(x) + h(y,f_1(x),\ldots,f_m(x)) \colon y \epsilon F\}$$

$$\begin{aligned} &= f_o(x) && \text{if } f_i(x) \leq 0, \quad i=1,2,\ldots,m. \\ &\ \ \infty && \text{if } f_i(x) > 0 \quad \text{for some } i. \end{aligned}$$

$$\beta(y) = \inf_x\{f_o(x) + h(y,f_1(x),\ldots,f_m(x)) \colon x \epsilon E^n\} \quad \text{for all } y \epsilon F$$

The problems P and D become,

$$\text{P:} \quad \text{minimize}_x\{f_o(x) \colon f_i(x) \leq 0, \quad i=1,2,\ldots,m\}$$

$$\text{D: maximize}_y\{\inf_x\{f_o(x) + h(y,f_1(x),\ldots,f_m(x)) \colon x \epsilon E^n\} \colon y \epsilon F\}.$$

---

† $\lim_{k\to\infty} y^k = \infty$ iff $\lim_{k\to\infty} y_i^k = \infty$ for each $i \epsilon \{1,2,\ldots,r\}$.

It should be noted that problem P is the usual form of a nonlinear program. Under certain restriction of $f_o, f_1, f_2, \ldots, f_m$, and h the optimal solutions of problems P and D become equivalent, and hence it suffices to solve problem D. The solution of problem D corresponds to the known method of exterior point unconstrained minimization technique. The function h may be viewed as a penalty term. It should also be noted that condition (i) above states that no penalty is assigned in case of a feasible solution. Condition (ii) indicates that if x strays too far from the feasible region, the penalty term $h(y^k, f_1(x), \ldots, f_m(x))$ becomes too large, and as $y^k \to \infty$, the tendency would be to draw the unconstrained minimum towards the feasible region. The movement is thus from the infeasible towards the feasible region.

The following gives an outline of the steps for solving problem P above via the penalty function method.

(a) For some $y^1 \epsilon F$, find the unconstrained minimum of $\phi(.,y^1)$ denoted by x^1.

(b) Starting from x^1, find an unconstrained minimum of $\phi(.,y^2)$, where $y^2 > y^1$. The minimum is x^2.

(c) Proceeding in this fashion, minimizing $\phi(.,y^k)$ for an increasing sequence $\{y^k\}$ such that $\lim_{k\to\infty} y^k = \infty$, then under certain conditions, the generated sequence of minima $\{x^k\}$ converges to the optimal solution of the primal problem.

For more details and proofs one may refer to [21] and [58].

Different choices of the function h (satisfying conditions (i) and (ii) above) and the set F lead to different forms of penalty functions. The following choices of h and F give some examples of different penalty functions.

(i) $h(y,f_1(x),\ldots,f_m(x)) = \sum_{i=1}^{m} y_i^2 \max(f_i(x),0)$, and $F = E^m$.

(ii) $h(y,f_1(x),\ldots,f_m(x)) = y \sum_{i=1}^{m} [\min(0,-f_i(x))]^2$, and $F = E_+^1$.

(iii) $h(y,f_1(x),\ldots,f_m(x)) = \max\{y_1 f_1(x),\ldots,y_m f_m(x),0\}$, and $F = E_+^m$.

In closing this chapter, it is worthwhile mentioning that duality may be used as a stopping rule of some solution procedures. As discussed in detail in Chapter III, the following two problems P and D,

$$\text{P: minimize}_x\{\alpha(x)\colon x\epsilon E\}$$

$$\text{D: maximize}_y\{\beta(y)\colon y\epsilon F\}$$

where $\alpha\colon E\to E^1$, $\beta\colon F\to E^1$, $E\subset E^n$ and $y\subset E^r$, are said to be dual programs iff $\inf_x\{\alpha(x)\colon x\epsilon E\} = \sup_y\{\beta(y)\colon y\epsilon F\}$. At any stage of a solution procedure, suppose we have $x\epsilon E$ and $y\epsilon F$ such that $\alpha(x) < \beta(y) + \epsilon$, for some sufficiently small $\epsilon > 0$, then we may stop. See, for example, [58].

CHAPTER II

OPTIMALITY CONDITIONS IN NONLINEAR PROGRAMMING

We have discussed in the previous chapter the important role that optimality conditions play in nonlinear programming both from the theoretical and computational points of view. The best known necessary optimality criteria when the functions are differentiable are the Fritz John and the Kuhn-Tucker conditions [27], [30]. In Section 3 later, we discuss these conditions and the generalizations that have been made so far by others. The purpose of this chapter is to investigate the necessary and sufficient conditions for optimality when the functions involved are nondifferentiable. Further, the convexity assumption is replaced by supportability and continuity assumptions of the functions involved are also relaxed in some cases.

In order to state and prove the various conditions for optimality in Section 3 of this chapter, we need several definitions and theorems. As mentioned earlier in Chapter I, we replace the concept of a gradient vector associated with a differentiable function by the known concept of a subgradient. We approach this, however, via the concept of outernormals to a set which is discussed in detail in the next section. Closely related to the existence of a nonzero outernormal to the set above or below a given function is the concept of supportability of the function. This notion permits us to relax the convexity (and continuity) assumptions in many theorems later. The concept of supportability along with various forms of convexity such as local convexity, quasi-convexity and pseudo-convexity are also discussed in the next section. In Section 2, we define and examine the properties of various cones corresponding to a given set. We present some new results which along with some existing results are used to prove the relationship between an outernormal to an intersection of sets and outernormals to the individual sets. This result is frequently used in Section 3 for deriving the conditions for optimality. In Section 3 we first consider a program where the functions involved are convex[†] (or locally convex) but not necessarily differentiable. Using the saddle value theorem, the Kuhn-Tucker conditions are then extended to non-differentiable functions. We then relax the differentiability assumption of the functions involved and assume that they are locally supportable from either above or below. This leads to an extension of the Fritz John conditions. Imposing some other additional restrictions gives us a further generalization of the Kuhn-Tucker conditions. We then investigate a necessary and sufficient condition for an interior point to be optimal. Further we develop some sufficient conditions by relaxing all restrictions concerning the constraint functions.

1. Outernormals and Subgradients

As mentioned earlier, the purpose of this section is to define and discuss the concepts of subgradients, supportable functions, and

---

[†]It is well known that a convex function which is defined on a convex set is continuous on the interior of the domain. Furthermore, the function is differentiable almost everywhere.

various forms of convexity. This is done using outernormals to sets and their associated supporting hyperplanes. We first give the following definition of a hyperplane and then consider the notion of a supporting hyperplane of a set at a boundary point.

Definition 1

A nonempty set H in E^n is said to be a *hyperplane through* $\bar{x}\epsilon E^n$ iff for some nonzero $p\epsilon E^n$, $H = \{x\colon x\epsilon E^n, \langle p,x-\bar{x}\rangle = 0\}$.

A hyperplane divides the space E^n into two halfspaces $H^- = \{x\colon x\epsilon E^n, \langle p,x-\bar{x}\rangle \leq 0\}$ and $H^+ = \{x\colon x\epsilon E^n, \langle p,x-\bar{x}\rangle \geq 0\}$.

Definition 2

Let A be a nonempty set in E^n and $\bar{x}\epsilon\partial A$. A hyperplane H in E^n is said to be a *supporting hyperplane of A at* $\bar{x}$ iff $\bar{x}\epsilon H$ and $A\subset H^-$ or $A\subset H^+$. A is said to be *supportable at* $\bar{x}$ iff there exists a supporting hyperplane of A at $\bar{x}$. The hyperplane H is said to *locally support A at* $\bar{x}$ iff $\bar{x}\epsilon H$ and for some neighborhood N about $\bar{x}$, $A\cap N\subset H^-$ (or $A\cap N\subset H^+$). A is said to be *locally supportable at* $\bar{x}$ iff there exists a locally supporting hyperplane of A at $\bar{x}$.

It can be shown that every nonempty convex set is supportable at every boundary point. See, for example, Valentine [51].

Definition 3

Let A be a nonempty set in E^n and $\bar{x}\epsilon\bar{A}$. A vector $p\epsilon E^n$ is said to be an *outernormal to A at* $\bar{x}$ iff $\langle p,x-\bar{x}\rangle \leq 0$ for all $x\epsilon\bar{A}$.

Outernormals are abbreviated to *o.n.* throughout the study. The following assertions follow immediately from the above definition.

Remark 1

(i) The zero vector is an *o.n.* to any nonempty set, at any point belonging to the closure of the set. Moreover, if $\bar{x} \in \text{int}(A)$, then the only *o.n.* to A at $\bar{x}$ is the zero vector.

(ii) Let A and B be nonempty sets in E^n. Let $A \subset B$ and $\bar{x} \in \bar{A} \cap \bar{B}$. If $p \in E^n$ is an *o.n.* to B at $\bar{x}$, then it also is an *o.n.* to A at $\bar{x}$. The converse is not necessarily true.

(iii) If p^1 and $p^2 \in E^n$ are both outernormals to a nonempty set A at $\bar{x} \in \bar{A}$, then $\lambda p^1 + \mu p^2$ is an *o.n.* to A at $\bar{x}$, where λ and μ are arbitrary nonnegative scalars.

As seen from the above remark, the only *o.n.* to a set at an interior point is the zero vector. This case is of no practical value and the interesting cases arise when the point under consideration is a boundary point. Definition 3 above includes both cases, however, since this is found more convenient in dealing with cones and their dual cones in Section 2.

Sometimes one may be interested in considering points that belong to a set in a small neighborhood of a given boundary point, e.g. in the theory of local optima. It may be the case that no nonzero outernormals to the set exist at that point, whereas such outernormals do exist if a small neighborhood about that point is considered. This motivates the following definition of a local outernormal.

Definition 4

Let A be a nonempty set in E^n and $\bar{x} \in \partial A$. A vector $p \in E^n$ is said to be a *local outernormal* to A at $\bar{x}$ iff there exists a neighborhood N about $\bar{x}$ such that p is an *o.n.* to $A \cap N$ at $\bar{x}$.

It follows immediately from the definition that if p is an *o.n.* to a nonempty set A at $\bar{x} \in \partial A$, then p is a local outernormal to A at $\bar{x}$, but not conversely. A local outernormal is abbreviated by *l.o.n.* in this presentation.

Definition 5

Let A be a nonempty set in E^n, then the *convex hull of A*, denoted by [A], is the minimum convex set that contains A.

Clearly if A is a convex set, then [A] = A. It can be shown that $x \in [A]$ iff $x = \sum_{i=1}^{n+1} \mu_i x^i$, where $x^i \in A$, $\mu_i \geq 0$ for $i=1,2,\ldots,n+1$ and $\sum_{i=1}^{n+1} \mu_i = 1$. See, for example, Valentine [51].

The following remark shows the close relationship between the notions of supportability, outernormals, and convex hulls.

Remark 2

Let A be a nonempty set in E^n and let $\bar{x} \in \partial A$. The following assertions are equivalent.

(i) A is supportable (locally supportable) at $\bar{x}$.

(ii) $\bar{x} \in \partial [A]$. ($\bar{x} \in \partial [A \cap N]$ for some neighborhood N about $\bar{x}$).

(iii) There exists a nonzero *o.n.* (*l.o.n.*) to A at $\bar{x}$.

Proof. We first show that (i) implies (ii), then show that (ii) implies (iii), and finally show that (iii) implies (i). First we assume that A is supportable at $\bar{x}$. Then by Definition 2 there exists a hyperplane H such that $\bar{x} \in H$ (and hence $\bar{x} \in \partial H^-$) and $A \subset H^-$. Since H^- is a convex set that contains A, then $[A] \subset H^-$. We assert that $\bar{x} \in \partial [A]$ because if not then $\bar{x} \in \text{int}([A])$, and hence $\bar{x} \in \text{int}(H^-)$, a contradiction. Therefore

$\bar{x}\epsilon\partial[A]$. We now assume that $\bar{x}\epsilon\partial[A]$. Since $[A]$ is a convex set, then there exists a supporting hyperplane of $[A]$ at $\bar{x}$. Let this hyperplane be $H = \{x: x\epsilon E^n, \langle p,x-\bar{x}\rangle = 0\}$, where p is a nonzero vector in E^n. It follows immediately that p is an *o.n.* to H^- at $\bar{x}$, and hence by Remark 1, p is an *o.n.* to both $[A]$ and A at $\bar{x}$. Finally we assume that there exists a nonzero vector $p\epsilon E^n$ such that $\langle p,x-\bar{x}\rangle \le 0$ for all $x\epsilon A$. We construct the hyperplane $H = \{x: \langle p,x-\bar{x}\rangle = 0, x\epsilon E^n\}$. It is immediate that this hyperplane supports A at $\bar{x}$. This completes the proof.

We now define a minimum (locally minimum) point of a set. Remark 3 that follows gives a necessary and sufficient condition for a point to be a minimum (locally minimum) point. These results are used later in Section 3.

Definition 6

Let A be a nonempty set in E^{n+1} of the following form,

$$A = \{(x,y): x\epsilon E\subset E^n,\ y\epsilon D(x)\subset E^1\},$$

where $D(x)$ is a subset of E^1 which is related to the point $x\epsilon E$. $(\bar{x},\bar{y})\epsilon A$ is said to be a *minimum point of A* iff $\bar{y} \le y$ for all $(x,y)\epsilon A$. A point (x^o,y^o) is said to be a *locally minimum point of A* iff there exists a neighborhood N about (x^o,y^o) such that (x^o,y^o) is a minimum point of $A\cap N$.

Similar definitions may be adopted for *maximum* and *locally maximum* points. As an example of a minimum point, we consider the set $A = \{(x,y): x\epsilon E,\ y \ge f(x)\}$. In this case, the set $D(x) = \{y: y \ge f(x)\}$.

It is obvious that $(\bar{x},f(\bar{x}))\epsilon A$ is a minimum point of A iff $f(x) \geq f(\bar{x})$ for all $x\epsilon E$.

The following remark gives a necessary and sufficient characterization of a minimum point (maximum point).

Remark 3

Let A be a nonempty set in E^{n+1} as in Definition 6. $(\bar{x},\bar{y})\epsilon A$ is a minimum point of A iff the n+1 dimensional vector $(0,0,\ldots,0,-1)$ is an *o.n.* to A at $(\bar{x},\bar{y})$. Similarly $(\hat{x},\hat{y})\epsilon A$ is a maximum point of A iff the n+1 dimensional vector $(0,0,\ldots,0,1)$ is an *o.n.* to A at $(\hat{x},\hat{y})$.

Proof. First assume that $(0,0,\ldots,0,-1)$ is an *o.n.* to A at $(\bar{x},\bar{y})$. Then by Definition 3, $\langle(0,0,\ldots,0,-1), (x,y) - (\bar{x},\bar{y})\rangle \leq 0$ for all $(x,y)\epsilon A$. This implies that $y \geq \bar{y}$ for all $(x,y)\epsilon A$ and hence $(\bar{x},\bar{y})$ is a minimum point of A. Conversely, if $(\bar{x},\bar{y})$ is a minimum point of A, then $y \geq \bar{y}$ for all $(x,y)\epsilon A$, and hence $(0,0,\ldots,0,-1)$ is an *o.n.* to A at $(\bar{x},\bar{y})$.

A similar argument proves the assertion about maximum points. It may be noted that similar characterization of a local minimum (local maximum) point can be obtained by replacing the notion of an *o.n.* in Remark 3 above by that of a *l.o.n.*

We now define the epigraph and the hypograph of a function. These definitions are used in turn in defining the supportability and subgradients of the function. As mentioned previously these two concepts play an important role in this study.

Definition 7

Consider a function $f\colon E^n \to E^1$. The set A is said to be the *epigraph of f* iff $A = \{(x,y)\colon x \in E^n,\ y \in E^1,\ y \geq f(x)\}$. The set B is said to be the *hypograph of f* iff $B = \{(x,y)\colon x \in E^n,\ y \in E^1,\ y \leq f(x)\}$.

It should be noted that in the above definition we have placed no restrictions of any kind on the nature of the function f. We now consider the following definition of supportable functions which is followed by two theorems that give different equivalent necessary and sufficient conditions for supportability.

Definition 8

A function $f\colon E^n \to E^1$ is said to be *supportable from below (supportable from above)* at $\bar{x} \in E^n$ iff there exists a hyperplane $H \subset E^{n+1}$ which supports the epigraph (hypograph) of f at $(\bar{x}, f(\bar{x}))$. f is said to be *locally supportable from below (locally supportable from above)* at $\bar{x} \in E^n$ iff there exists a hyperplane that locally supports the epigraph (hypograph) of f at $(\bar{x}, f(\bar{x}))$.

Theorem 1

Consider the function $f\colon E^n \to E^1$ and let A be its epigraph. f is supportable (locally supportable) from below at $\bar{x} \in E^n$ iff any one of the following equivalent conditions hold.

(i) $(\bar{x}, f(\bar{x})) \in \partial[A]$ $\{\in \partial[A \cap N]$ for some neighborhood N about $(\bar{x}, f(\bar{x}))\}$.

(ii) There exists a nonzero *o.n.* *(l.o.n.)* to A at $(\bar{x}, f(\bar{x}))$.

(iii) There exists a vector $\theta(\bar{x}) \in E^n$ such that, $f(x) \geq f(\bar{x}) + \langle \theta(\bar{x}), x - \bar{x} \rangle$ for all $x \in E^n$ (for all x in some neighborhood N about $\bar{x}$).

Proof. From Definition 8 and Remark 2 it is immediate that both conditions (i) and (ii) are both necessary and sufficient for supportability of f at $\bar{x}$. Hence it suffices to show that conditions (ii) and (iii) above are equivalent. We first show that condition (ii) implies condition (iii). Assume that $(p,p_{n+1}) \in E^{n+1}$ is a non-zero *o.n.* to A at $(\bar{x},f(\bar{x}))$. This implies that,

$$<(p,p_{n+1}),(x,y) - (\bar{x},f(\bar{x}))> \leq 0 \qquad \text{for all } x\epsilon E^n,\ y \geq f(x) \tag{1}$$

We first show that $p_{n+1} < 0$ by showing that it can neither be equal to zero, nor greater than zero. Suppose on the contrary that $p_{n+1} = 0$, then for some $j\epsilon\{1,2,\ldots,n\}$, $p_j \neq 0$. Choose $(x,f(x))\epsilon A$ such that $x_i = \bar{x}_i$ for all $i \neq j$ and $x_j = \bar{x}_j + p_j$. Therefore, $<(p,p_{n+1}), \quad (x,f(x)) - (\bar{x},f(\bar{x}))> = \sum_{i=1}^{n} p_i(x_i-\bar{x}_i) = p_j^2 > 0$, which contradicts (1). Therefore $p_{n+1} \neq 0$. On the other hand, suppose that $p_{n+1} > 0$, then consider $(\bar{x},f(\bar{x}) + 1)\epsilon A$. Therefore, $<(p,p_{n+1}),(\bar{x},f(\bar{x})+1) - (\bar{x},f(\bar{x}))> = p_{n+1} > 0$, which again contradicts (1). Therefore, $p_{n+1} < 0$ as asserted above. By dividing inequality (1) by $|p_{n+1}|$, choosing $y = f(x)$, and denoting $p/|p_{n+1}|$ by $\theta(\bar{x})$, the following inequality is obtained.

$$<(\theta(\bar{x}),-1),(x,f(x)) - (\bar{x},f(\bar{x}))> \leq 0 \qquad \text{for all } x\epsilon E^n \tag{2}$$

By rearranging terms in the above inequality condition (iii) follows. To show the converse, assume condition (iii). Then

$$y \geq f(\bar{x}) + \langle\theta(\bar{x}),x-\bar{x}\rangle \qquad \text{for all } x\epsilon E^n, \text{ all } y \geq f(x). \qquad (3)$$

Rearranging the terms in the above inequality, Inequality (1) is obtained where $(p,p_{n+1}) = (\theta(\bar{x}),-1)$. This completes the proof.

Theorem 2

Consider the function f: $E^n \to E^1$ and let B be its hypograph. f is supportable (locally supportable) from above at $\bar{x}\epsilon E^n$ iff any one of the following equivalent conditions hold.

(i) $(\bar{x},f(\bar{x}))\epsilon\partial[B]$ $\{\epsilon\partial[B\cap N]$ for some neighborhood N about $(\bar{x},f(\bar{x}))\}$.

(ii) There exists a nonzero *o.n.* *(l.o.n.)* to B at $(\bar{x},f(\bar{x}))$.

(iii) There exists a vector $\theta(\bar{x})\epsilon E^n$ such that, $f(x) \leq f(\bar{x}) + \langle\theta(\bar{x}),x-\bar{x}\rangle$ for all $x\epsilon E^n$ (for all x in some neighborhood N about $\bar{x}$).

The proof is similar to that of Theorem 1 and is omitted.

We would now like to consider points in the domain of the function where it is not necessarily differentiable. Later we introduce the notion of a subgradient and show that if the function is differentiable at a given point, then it has a unique subgradient, namely the gradient vector which is defined below.

Definition 9

A function f: $E^n \to E^1$ is said to be *differentiable at* $x\epsilon E^n$ iff there exists a vector $\nabla f(x)\epsilon E^n$ called the *gradient of* f *at* x such that the function $h(x,.)$: $E^n \to E^1$ defined by

$$h(x,u) = (f(x+u) - f(x) - \langle \nabla f(x),u \rangle)/\|u\| \qquad u \in E^n,\ u \neq 0$$

tends to zero as $\|u\| \to 0$. f is said to be *differentiable* iff it is differentiable at all $x \in E^n$.

It can be shown that the gradient vector in the above definition is unique and is given by $\nabla f(x) = (\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n})$. See, for example, [5].

We now introduce the notion of a subgradient of a function at a point in its domain via the outernormals to the epigraph or the hypograph of the function. This definition is a generalization of the definition given by Moreau[†] [40] and adopted by Brønsted and Rockafeller [8], and others.

Definition 10

Consider the function $f: E^n \to E^1$. $\theta(\bar{x}) \in E^n$ is said to be a *subgradient of f at* $\bar{x}$ iff one of the following occurs.

(i) $\theta(\bar{x}) = \nabla f(\bar{x})$ if the latter exists.

(ii) $(\theta(\bar{x}),-1)$ is a *l.o.n.* to the epigraph of f at $(\bar{x},f(\bar{x}))$.

(iii) $(-\theta(\bar{x}),1)$ is a *l.o.n.* to the hypograph of f at $(\bar{x},f(\bar{x}))$.

It follows from Theorems 1 and 2 that case (ii) in the definition above refers to a locally supportable function from below and that case (iii) refers to a locally supportable function from above.

Sometimes we find it convenient to consider a translation of

[†]Let $f: E^n \to E^1$ be a convex function. $x^* \in E^n$ is said to be a subgradient of f at $\bar{x} \in E^n$ iff $f(x) \geq f(\bar{x}) + \langle x^*, x-\bar{x} \rangle$ for all $x \in E^n$.

the function f. In this case we may rewrite cases (ii) and (iii) of the above definition as follows. If $(\theta(\bar{x}),-1)$ is a *l.o.n.* to the set $\{(x,y)\colon x\epsilon E^n, y\epsilon E^1, y \geq f(x) - f(\bar{x})\}$ at $(\bar{x},o)$, then $\theta(\bar{x})$ is a subgradient of f at $\bar{x}$. Similarly, if $(-\theta(\bar{x}),1)$ is a *l.o.n.* to the set $\{(x,y)\colon x\epsilon E^n, y\epsilon E^1, y \leq f(x) - f(\bar{x})\}$ at $(\bar{x},o)$, then $\theta(\bar{x})$ is a subgradient of f at $\bar{x}$.

Theorem 3

Let $f\colon E^n \to E^1$ be differentiable at $\bar{x}\epsilon E^n$. Then $\nabla f(\bar{x})$ is the unique subgradient of f at $\bar{x}$.

Proof. We prove the result by showing that any arbitrary subgradient $\theta(\bar{x})$ is equal to the gradient vector. Let $\theta(\bar{x})\epsilon E^n$ be a subgradient of f at $\bar{x}$. If case (i) of Definition 10 is encountered, then $\theta(\bar{x}) = \nabla f(\bar{x})$. We now consider case (ii) of Definition 10, where $(\theta(\bar{x}),-1)$ is a *l.o.n.* of the epigraph of f at $(\bar{x},f(\bar{x}))$. By Theorem 1 we conclude that,

$$f(x) \geq f(\bar{x}) + \langle\theta(\bar{x}),x-\bar{x}\rangle \qquad \text{for all } x\epsilon N_1 \tag{4}$$

where N_1 is some neighborhood about $\bar{x}$. But since f is differentiable at $\bar{x}$, then from Definition 9, we conclude that,

$$f(x) = f(\bar{x}) + \langle\nabla f(\bar{x}),x-\bar{x}\rangle + 0\|x-\bar{x}\| \qquad \text{for all } x\epsilon E^n \tag{5}$$

where $0\|x-\bar{x}\|$ satisfies the property that $\lim_{\|x-\bar{x}\|\to 0} 0\|x-\bar{x}\|/\|x-\bar{x}\| = 0$. Combining (4) and (5) we conclude that,

$$<\nabla f(\bar{x}) - \theta(\bar{x}), x-\bar{x}> + 0\|x-\bar{x}\| \geq 0 \qquad \text{for all } x\epsilon N_1$$

We assert that $\nabla f(\bar{x}) = \theta(\bar{x})$. If $\nabla f_j(\bar{x}) \neq \theta_j(\bar{x})$ for some $j\epsilon\{1,2,\ldots,n\}$ then consider x^δ given by $x_i^\delta = \bar{x}_i$ for all $i \neq j$ and $x_j^\delta = \bar{x}_j - \delta(\nabla_j f(\bar{x}) - \theta_j(\bar{x}))$, where $\delta > 0$. It can be easily shown that there exists an $\bar{\delta} > 0$ such that $x^\delta \epsilon N_1$ for all $\delta\epsilon[0,\bar{\delta}]$. Therefore, letting $x = x^\delta$ it follows that for all $\delta\epsilon[0,\bar{\delta}]$,

$$-\delta(\nabla_j f(\bar{x}) - \theta_j(\bar{x}))^2 + 0\|(0,0,\ldots,-\delta(\nabla_j f(x) - \theta_j(\bar{x})),0,\ldots,0)\| \geq 0 \quad (6)$$

By dividing (6) by $\|x^\delta-\bar{x}\| = \delta|\nabla_j f(\bar{x})-\theta_j(\bar{x})|$ and taking the limit as $\delta\to 0$ we arrive to an immediate contradiction. Therefore $\nabla f(\bar{x}) = \theta(\bar{x})$. A similar argument shows that $\theta(\bar{x}) = \nabla f(\bar{x})$ when case (iii) of Definition 10 is encountered. This completes the proof.

We now discuss various forms of convexity including quasi-convexity and pseudo-convexity. Since the functions are not necessarily differentiable, the notion of a subgradient is used whenever necessary. Furthermore, since the concept of supportability relaxes the concept of convexity in discussing optimality conditions, the relationship between these two concepts is also discussed.

Definition 11

Consider the function $f: E^n\to E^1$. f is said to be *convex* iff for every x^1 and $x^2\epsilon E^n$, $f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda)f(x^2)$ for all $\lambda\epsilon(0,1)$. f is said to be *concave* iff $-f$ is convex.

Definition 12

Consider the function f: $E^n \to E^1$. f is said to be *locally convex at* $\bar{x} \epsilon E^n$ iff there exists a neighborhood N about $\bar{x}$ such that f is convex on N, i.e. for every x^1 and $x^2 \epsilon N$, $f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda)f(x^2)$ for all $\lambda \epsilon (0,1)$. f is said to be *locally concave at* $\bar{x} \epsilon E^n$ iff -f is locally convex at $\bar{x}$.

f is said to be *strictly (convex, concave, locally convex, locally concave)* iff the inequalities in the above two definitions hold strictly for $x^1 \neq x^2$ for all $\lambda \epsilon (0,1)$.

We now give different characterizations of convexity. We show that a function is convex iff it is supportable everywhere. We also give characterization of a convex function via its subgradients and its epigraph.

Theorem 4

Consider the function f: $E^n \to E^1$. The following assertions are equivalent.

(i) f is a convex (concave) function.

(ii) The epigraph (hypograph) of f is a convex set.

(iii) f is supportable from below (above) everywhere.

(iv) For all x^1 and $x^2 \epsilon E^n$, $f(x^2) \geq f(x^1) + \langle \theta(x^1), x^2 - x^1 \rangle$ where $\theta(x^1)$ is a subgradient of f at x^1. ($f(x^2) \leq f(x^1) + \langle \theta(x^1), x^2 - x^1 \rangle$).

Proof. We prove the result by showing that (i) implies (ii), (ii) implies (iii), (iii) implies (iv), and finally (iv) implies (i). We first assume that f is a convex function, and furthermore assume that $(x^1, y^1) \epsilon A$ and $(x^2, y^2) \epsilon A$, where A is the epigraph of f.

By convexity of f we conclude that,

$$\lambda y^1 + (1-\lambda)y^2 \geq \lambda f(x^1) + (1-\lambda)f(x^2) \geq f(\lambda x^1 + (1-\lambda)x^2)$$

$$\text{for all } \lambda\epsilon(0,1).$$

This implies that $\lambda(x^1,y^1) + (1-\lambda)(x^2,y^2)\epsilon A$ for all $\lambda\epsilon(0,1)$, which further implies convexity of A. The assertions that (ii) implies (iii) and that (iii) implies (iv) follow immediately from Theorem 1. We finally show that (iv) implies (i). Assume that x^1 and $x^2 \epsilon E^n$ and consider $x_\lambda = \lambda x^1 + (1-\lambda)x^2$ for some $\lambda\epsilon(0,1)$. Therefore by hypothesis,

$$f(x^1) \geq f(x_\lambda) + \langle\theta(x_\lambda),x^1-x_\lambda\rangle = f(x_\lambda) + (1-\lambda)\langle\theta(x_\lambda),x^1-x^2\rangle \quad (7)$$

$$f(x^2) \geq f(x_\lambda) + \langle\theta(x_\lambda),x^2-x_\lambda\rangle = f(x_\lambda) + \lambda\langle\theta(x_\lambda),x^2-x^1\rangle \quad (8)$$

where $\theta(x_\lambda)$ is a subgradient of f at x_λ. Multiplying (7) by λ and (8) by $(1-\lambda)$ and adding the two inequalities we obtain the following inequality.

$$\lambda f(x^1) + (1-\lambda)f(x^2) \geq f(x_\lambda) = f(\lambda x^1 + (1-\lambda)x^2).$$

But since this is true for all $\lambda\epsilon(0,1)$, then f is a convex function. This completes the proof.

Similar results can be obtained in the case of locally convex (locally concave) functions. We now consider the following remark which gives some results concerning subgradients of convex functions.

Remark 4

Let $f: E^n \to E^1$ be a convex function. Then,

(i) $<\theta(x^1) - \theta(x^2), x^1 - x^2> \geq 0$ for any x^1 and $x^2 \epsilon E^n$, where $\theta(x^1)$ and $\theta(x^2)$ are subgradients of f at x^1 and x^2.

(ii) $<\theta(\lambda x^1 + (1-\lambda)x^2), x^2 - x^1>$ is a non-increasing function of λ, for all $\lambda \epsilon [0,1]$, where x^1 and x^2 are fixed points in E^n, and $\theta(\lambda x^1 + (1-\lambda)x^2)$ is a subgradient of f at $\lambda x^1 + (1-\lambda)x^2$.

Proof. By Theorem 4, the following two inequalities are obtained. $f(x^2) \geq f(x^1) + <\theta(x^1), x^2 - x^1>$ and $f(x^1) \geq f(x^2) + <\theta(x^2), x^1 - x^2>$.

Adding the above two inequalities, the result of (i) is immediate. We now choose λ_1 and λ_2 such that $0 \leq \lambda_1 < \lambda_2 \leq 1$. Let $x(\lambda_1) = \lambda_1 x^1 + (1-\lambda_1)x^2$ and $x(\lambda_2) = \lambda_2 x^1 + (1-\lambda_2)x^2$, and hence $x(\lambda_2) - x(\lambda_1) = (\lambda_2 - \lambda_1)(x^1 - x^2)$. We apply the result of (i) which implies that $<\theta(x(\lambda_1)) - \theta(x(\lambda_2)), x(\lambda_1) - x(\lambda_2)> \geq 0$. In other words, $(\lambda_2 - \lambda_1)<\theta(\lambda_1 x^1 + (1-\lambda_1)x^2) - \theta(\lambda_2 x^1 + (1-\lambda_2)x^2), x^2 - x^1> \geq 0$. Since $\lambda_2 - \lambda_1 > 0$, the result become immediate by dividing the above inequality by $\lambda_2 - \lambda_1$.

We now consider quasi-convex and pseudo-convex functions. These functions enjoy some, but not all, of the properties of convex functions.

Definition 13

$f: E^n \to E^1$ is said to be *quasi-convex* iff any of the following two equivalent conditions† is satisfied.

---

†The equivalence of the two definitions may be found in [35].

(i) Given any $\bar{x}\epsilon E^n$, if $f(x) \leq f(\bar{x})$ then $f(\lambda x + (1-\lambda)\bar{x}) \leq f(\bar{x})$ for all $\lambda\epsilon(0,1)$.

(ii) Given any $\delta\epsilon E^1$, then $A(\delta) = \{x: x\epsilon E^n, f(x) \leq \delta\}$ is a convex set.

It can be shown [35] that if f is differentiable, then it is quasi-convex iff for any x^1 and $x^2 \epsilon E^n$ such that $f(x^1) \leq f(x^2)$, the inequality $<\nabla f(x^2), x^1 - x^2> \leq 0$ holds.

f is said to be *quasi-concave* iff -f is quasi-convex. It may be noted that the sum of two quasi-convex (quasi-concave) functions is not necessarily quasi-convex (quasi-concave).

Definition 14

$f: E^n \to E^1$ is said to be *strictly quasi-convex*† iff $x \neq \bar{x}$ and $f(x) \leq f(\bar{x})$ implies that $f(\lambda x + (1-\lambda)\bar{x}) < f(\bar{x})$ for all $\lambda\epsilon(0,1)$. $f: E^n \to E^1$ is said to be *s-strictly quasi-convex*†† iff $f(x) < f(\bar{x})$ implies that $f(\lambda x + (1-\lambda)\bar{x}) < f(\bar{x})$ for all $\lambda\epsilon(0,1)$. f is said to be *strictly quasi-concave* iff -f is strictly quasi-convex. f is said to be *s-strictly quasi-concave* iff -f is *s*-strictly quasi-convex.

The following remark shows that the maximum of a quasi-convex function and the zero function is quasi-convex.

Remark 5

Let $f: E^n \to E^1$ be quasi-convex. Then $g = \max(0,f)$ is also

---

†Roode [45] refers to such a function as a monotonic strictly quasi-convex function. By our definition, strict quasi-convexity implies quasi-convexity.

††Karamadian [28], Mangasarian [34], and Roode [45], refer to such a function as a strictly quasi-convex function. An *s*-strictly quasi-convex function is not necessarily quasi-convex.

quasi-convex.

Proof. Consider x^1 and $x^2 \in E^n$ such that $g(x^1) \le g(x^2)$, i.e. $\max(0,f(x^1)) \le \max(0,f(x^2))$. This implies that either both $f(x^1)$ and $f(x^2) \le 0$, or $f(x^2) \ge f(x^1)$ and $f(x^2) > 0$. We consider both cases as follows.

(i) If $f(x^1) \le 0$ and $f(x^2) \le 0$, then $g(x^1) = g(x^2) = 0$. But $g(\lambda x^1 + (1-\lambda)x^2) = \max(0,f(\lambda x^1 + (1-\lambda)x^2)) = 0$ for all $\lambda\epsilon(0,1)$ by quasi-convexity of f.

(ii) If $f(x^2) > 0$, then $f(\lambda x^1 + (1-\lambda)x^2)$ is either positive or nonpositive. In the former case, $g(\lambda x^1 + (1-\lambda)x^2) = f(\lambda x^1 + (1-\lambda)x^2) \le f(x^2) = g(x^2)$. In the latter case, $g(\lambda x^1 + (1-\lambda)x^2) = 0 < f(x^2) = g(x^2)$.

At any rate $g(\lambda x^1 + (1-\lambda)x^2) \le g(x^2)$ for all $\lambda\epsilon(0,1)$, and the proof is complete.

We now consider pseudo-convex functions introduced by Mangasarion [34]. Mangasarian considered only differentiable functions. We relax the differentiability assumption by considering subgradients.

Definition 15

$f\colon E^n \to E^1$ is said to be *pseudo-convex* iff for any given $\bar{x}\epsilon E^n$, $\langle\theta(\bar{x}),x-\bar{x}\rangle \ge 0$ implies that $f(x) \ge f(\bar{x})$ where $\theta(\bar{x}) \in E^n$ is a subgradient of f at $\bar{x}$. f is said to be *pseudo-concave* iff -f is pseudo-convex.

It should be noted that every convex function is both quasi-convex and pseudo-convex. This follows immediately from Definitions 13 and 15 and Theorem 4. Obviously, the converse is not necessarily true. We now show that every pseudo-convex function is quasi-convex. Mangasarian [34] showed that every pseudo-convex function is *s*-strictly quasi-convex, when the function under consideration is differentiable.

Theorem 5

If f: $E^n \to E^1$ is pseudo-convex then it is quasi-convex.

Proof. Assume on the contrary that f is not quasi-convex. This implies that there exist x^1 and $x^2 \in E^n$ such that $f(x^1) \leq f(x^2)$ and $f(\bar{\lambda}x^1 + (1-\bar{\lambda})x^2) > f(x^2)$ for some $\bar{\lambda} \in (0,1)$. Then $f(\bar{x}) > f(x^2) \geq f(x^1)$, where $\bar{x} = \bar{\lambda}x^1 + (1-\bar{\lambda})x^2$. Let $\theta(\bar{x})$ be a subgradient of f at $\bar{x}$, and consider $<\theta(\bar{x}), x^2-\bar{x}>$. If $<\theta(\bar{x}), x^2-\bar{x}> \geq 0$ then $f(x^2) \geq f(\bar{x})$, a contradiction. If on the other hand $<\theta(\bar{x}), x^2-\bar{x}> < 0$, it follows that $<\theta(\bar{x}), x^1-\bar{x}> > 0$, and hence $f(x^1) \geq f(\bar{x})$, which is again a contradiction. This implies that f is quasi-convex and the proof is complete.

The converse of this theorem is obviously not true. An example given by Mangasarian [34] is $f(x) = x^3$ for all $x \in E^1$. f is quasi-convex but not pseudo-convex since it violates Definition 15 at $x = 0$.

Earlier, we showed that a convex function is supportable everywhere. At this stage it may be helpful to point out that one cannot draw any general relationships between supportability, quasi-convexity and pseudo-convexity. We now give two examples which show that supportability does not imply and is not implied by either quasi-convexity or pseudo-convexity. Consider the function f: $E^1 \to E^1$ defined by $f(x) = x^3 + x$ for all $x \in E^1$. It can easily be checked that f is both quasi-convex and pseudo-convex but is supportable neither from below nor from above at $x = 0$. On the other hand, the function f: $E^1 \to E^1$, defined by

$$f(x) = 1 \qquad \text{for all } x\epsilon[-1,1]$$

$$0 \qquad \text{for } x > 1 \text{ or } x < -1$$

is supportable everywhere either from below or above. However, f is neither quasi-convex nor pseudo-convex.

The results developed in this section concerning outernormals and subgradients are used throughout the study. We now make use of these results to give an important relationship between subgradients of two locally convex functions and subgradient of the sum of the functions. Starting with Theorem 4 as a definition of a subgradient of a convex function and by making use of the conjugate function theory discussed in Chapter III, Rockafeller [43] has established a similar result.

Theorem 6

Let f_1, f_2, and $f\colon E^n \to E^1$ be locally convex functions at $\bar{x}\epsilon E^n$, and $f(x) = f_1(x) + f_2(x)$ for all $x\epsilon N$, where N is some neighborhood about $\bar{x}$. $\theta(\bar{x})\epsilon E^n$ is a subgradient of f at $\bar{x}$ iff $\theta(\bar{x}) = \theta^1(\bar{x}) + \theta^2(\bar{x})$ where $\theta^1(\bar{x})$ and $\theta^2(\bar{x})$ are subgradients of f_1 and f_2 at $\bar{x}$.

Proof. We first assume that $\theta^1(\bar{x})$ and $\theta^2(\bar{x})$ are subgradients of f_1 and f_2 at $\bar{x}$. Therefore by Theorem 4 we conclude that,

$$f_1(x) \geq f_1(\bar{x}) + \langle\theta^1(\bar{x}), x-\bar{x}\rangle \quad \text{for all } x\epsilon N_1, \text{ and}$$

$$f_2(x) \geq f_2(\bar{x}) + \langle\theta^2(\bar{x}), x-\bar{x}\rangle \quad \text{for all } x\epsilon N_2$$

where N_1 and N_2 are neighborhoods of $\bar{x}$. By adding the above two inequalities, letting $\bar{N}$ be the minimum of N, N_1, and N_2, and denoting $\theta^1(\bar{x}) + \theta^2(\bar{x})$ by $\theta(\bar{x})$ we conclude that,

$$f(x) = f_1(x) + f_2(x) \geq f(\bar{x}) + \langle\theta(\bar{x}),x-\bar{x}\rangle \quad \text{for all } x\epsilon\bar{N}. \tag{9}$$

By Theorem 4 it follows that $\theta(\bar{x})$ is a subgradient of f at $\bar{x}$. Conversely, we assume that $\theta(\bar{x})$ is a subgradient of f at $\bar{x}$. Therefore, $f(x) \geq f(\bar{x}) + \langle\theta(\bar{x}),x-\bar{x}\rangle$ for all $x\epsilon N_3$, where N_3 is some neighborhood about $\bar{x}$. Since f_1, f_2, and f are locally convex at $\bar{x}$, then there exists a neighborhood N_4 about $\bar{x}$ such that f_1, f_2, and f are convex on N_4. Let $\tilde{N}$ be the minimum of N_3, N_4, and N and consider the two sets A and B in E^{n+1}.

$$A = \{(x,y)\colon x\epsilon\tilde{N}\subset E^n,\ y\epsilon E^1,\ y \geq f_1(x) - \langle\theta(\bar{x}),x-\bar{x}\rangle - f(\bar{x})\}.$$

$$B = \{(x,y)\colon x\epsilon\tilde{N}\subset E^n,\ y\epsilon E^1,\ y < -f_2(x)\}.$$

By Theorem 4, it follows that A and B are convex sets. We also show that A and B are disjoint. Assume on the contrary that $(x,y)\epsilon A\cap B$. Therefore,

$$y \geq f_1(x) - \langle\theta(\bar{x}),x-x\rangle - f(\bar{x}) \quad \text{and}$$

$$-y > f_2(x).$$

By adding the above two inequalities, it follows that $f(x) < f(\bar{x}) + \langle\theta(\bar{x}),x-\bar{x}\rangle$, which is a contradiction. This implies that $A \cap B = \phi$, and since A and B are convex, then there exists a hyperplane H that separates them. See, for example, [15] or [51]. We assert that $(\bar{x},-f_2(\bar{x})) \in H$ because if not then $(\bar{x},-f_2(\bar{x}))$ belongs to either $\mathrm{int}(H^-)$ or $\mathrm{int}(H^+)$. Therefore, there exists a neighborhood $N(\bar{x})$ about $\bar{x}$ such that either $N(\bar{x}) \subset \mathrm{int}(H^-)$ or $N(\bar{x}) \subset \mathrm{int}(H^+)$. It can easily be checked that $(\bar{x},-f_2(\bar{x})) \in (\partial A) \cap (\partial B)$ and hence $A \cap N(\bar{x}) \neq \phi$ and $B \cap N(\bar{x}) \neq \phi$. But since $A \subset H^+$ and $B \subset \mathrm{int}(H^-)$, then $H^+ \cap N(\bar{x}) \neq \phi$ and $H^- \cap N(\bar{x}) \neq \phi$ which violates the fact that either $N(\bar{x}) \subset \mathrm{int}(H^-)$ or $N(\bar{x}) \subset \mathrm{int}(H^+)$. Therefore $(\bar{x},-f_2(\bar{x})) \in H$, and hence there exists a non-zero vector $p = (\bar{p},p_{n+1}) \in E^{n+1}$ such that $H = \{(x,y)\colon x \in E^n,\ y \in E^1,\ \langle(p,p_{n+1}),(x,y) - (\bar{x},-f_2(\bar{x})\rangle = 0\}$. It should be noted that $\langle(\bar{p},p_{n+1}),(x,y) - (\bar{x},-f_2(\bar{x}))\rangle \geq 0$ for all $(x,y) \in A$ and $\langle(\bar{p},p_{n+1}),(x,y) - (\bar{x},-f_2(\bar{x}))\rangle < 0$ for all $(x,y) \in B$. Using an argument similar to that of the proof of Theorem 1, it can be shown that $p_{n+1} > 0$. Dividing the above two inequalities by p_{n+1} and rearranging terms we obtain the following inequalities.

$$y \geq -f_2(\bar{x}) - \langle(\bar{p}/p_{n+1}),x-\bar{x}\rangle \quad \text{for all } x \in \tilde{N}, \tag{10}$$

$$\text{and all } y \geq f_1(x) - \langle\theta(\bar{x}),x-\bar{x}\rangle - f(\bar{x})$$

$$y < -f_2(\bar{x}) - \langle(\bar{p}/p_{n+1}),x-\bar{x}\rangle \quad \text{for all } x \in \tilde{N} \tag{11}$$

$$\text{and all } y < -f_2(x)$$

By letting $y = f_1(x) - \langle\theta(\bar{x}),x-\bar{x}\rangle - f(\bar{x})$ in (10) and noting that (11) implies (13) the following inequalities are obtained.

$$f_1(x) \geq f_1(\bar{x}) + \langle\theta(\bar{x}) - (\bar{p}/p_{n+1}),x-\bar{x}\rangle \quad \text{for all } x\epsilon\tilde{N} \tag{12}$$

$$f_2(x) \geq f_2(\bar{x}) + \langle(\bar{p}/p_{n+1}),x-\bar{x}\rangle \quad \text{for all } x\epsilon\tilde{N} \tag{13}$$

Inequalities (10) and (11) imply that $\theta(\bar{x}) - \bar{p}/p_{n+1}$ and $\bar{p}/p_{n+1}$ are subgradients of f_1 and f_2. By letting $\theta^1(\bar{x}) = \theta(\bar{x}) - \bar{p}/p_{n+1}$ and $\theta^2(\bar{x}) = \bar{p}/p_{n+1}$ the result is immediate.

Corollary

Let $f,f_1,f_2,\ldots,f_m$: $E^n \to E^1$ be locally convex at $\bar{x}\epsilon E^n$, and let $f = \sum_{i=1}^{m} f_i$ in a neighborhood of $\bar{x}$. Then $\theta(\bar{x})$ is a subgradient of f at $\bar{x}$ iff $\theta(\bar{x}) = \sum_{i=1}^{m} \theta^i(\bar{x})$ where $\theta^i(\bar{x})$ is a subgradient of f_i at $\bar{x}$.

2. Minimal Cones and Their Dual Cones

In this section we consider minimal cones, minimal convex cones, and dual cones of sets at a given point. We develop some results which relate outernormals to a set with outernormals to minimal cones and minimal convex cones of the set. Since a nonlinear programming problem can be entirely formulated in terms of the epigraphs and hypographs of the functions involved in the program, then the results of this section can be used in developing optimality criteria. This is done in Section 3.

We now consider the following definitions and preliminary results concerning cones.

Definition 16

A nonempty set C in E^n is said to be a *cone with vertex* $\bar{x}\in E^n$ iff for all $x\in C$, $\bar{x} + \lambda(x-\bar{x})\in C$ for all $\lambda\geq 0$.

The following remark shows that if the vertex of the cone is an interior point of the cone, then the cone in E^n.

Remark 6

If C is a cone in E^n with vertex $\bar{x}\in \mathrm{int}(C)$, then $C = E^n$.

Proof. Since $\bar{x} \in \mathrm{int}(C)$, then there exists a neighborhood N about $\bar{x}$ such that $N\subset \mathrm{int}(C)$. We now consider an arbitrary $x\in E^n$. Therefore, if $x\in N$, then $x\in C$. If on the other hand $x\notin N$, then since $\bar{x}\in \mathrm{int}(N)$, then $x(\mu) = \mu x + (1-\mu)\bar{x}$ belongs to N for some $\mu\in(0,1)$. But since C is a cone with vertex $\bar{x}$, then $\bar{x} + \lambda(x(\mu) - \bar{x}) \in C$ for all $\lambda \geq 0$. By choosing $\lambda = 1/\mu$, then it follows immediately that $x \in C$. This implies that $C = E^n$.

Definition 17

Let C be a cone in E^n with vertex $\bar{x}$. $\sim C$ is said to be the *image of C* iff $\sim C = \{\bar{x} + \lambda(x-\bar{x})\colon x\in C,\ \lambda\leq 0\}$.

Definition 18

Let C be a cone in E^n with vertex $\bar{x}$. C^* is said to be the *dual cone of C* iff $C^* = \{p\colon \langle p,x-\bar{x}\rangle \leq 0, \text{ all } x\in C\}$.

It can be immediately checked that C^* is indeed a cone with the zero vector as a vertex. It should be noted that C^* is the collection of outernormals to C at $\bar{x}$.

Definition 19

A cone C in E^n with vertex $\bar{x}$ is said to be a *convex cone with vertex* $\bar{x}$ iff x^1 and $x^2 \in C$ imply that $x^1 + x^2 - \bar{x} \in C$.

We point out that the above definition agrees with the known notion of convexity. We want to show that if x^1 and $x^2 \epsilon C$ then $\lambda x^1 + (1-\lambda)x^2 \epsilon C$ for all $\lambda \epsilon (0,1)$. Since $x^1 \epsilon C$, then $y^1 = \bar{x} + \lambda(x^1-\bar{x}) \epsilon C$ and since $x^2 \epsilon C$, then $y^2 = \bar{x} + (1-\lambda)(x^2-\bar{x}) \epsilon C$ for all $\lambda \epsilon (0,1)$. Therefore by Definition 19, $y^1 + y^2 - \bar{x} = \lambda x^1 + (1-\lambda)x^2 \epsilon C$ for all $\lambda \epsilon (0,1)$. Conversely, we assume that if x^1 and $x^2 \epsilon C$, then $\lambda x^1 + (1-\lambda)x^2 \epsilon C$ for all $\lambda \epsilon (0,1)$. We then show that $x^1 + x^2 - \bar{x} \epsilon C$. Letting $\lambda = 1/2$, then $1/2(x^1+x^2) \epsilon C$. But since C is a cone, then $\bar{x} + \mu(1/2(x^1+x^2) - \bar{x}) \epsilon C$ for all $\mu \geq 0$. The result is immediate by letting $\mu = 2$.

Remark 7

Let C be a cone in E^n with vertex $\bar{x}$. Then [C] is a convex cone with vertex $\bar{x}$.

Proof. We first show that [C] is a cone with vertex $\bar{x}$. Let $x \epsilon [C]$ then we need to show that $\bar{x} + \lambda(x-\bar{x}) \epsilon [C]$ for all $\lambda \geq 0$. Since $x \epsilon [C]$ then $x = \sum_{i=1}^{n+1} \mu_i x^i$ where $x^i \epsilon C$ and $\mu_i \geq 0$ $(i=1,2,\ldots,n+1)$ and $\sum_{i=1}^{n+1} \mu_i = 1$. But since C is a cone with vertex $\bar{x}$, then

$$\bar{x} + \lambda(x-\bar{x}) = \bar{x} + \lambda\left(\sum_{i=1}^{n+1} \mu_i x^i - \bar{x}\right) = \sum_{i=1}^{n+1} \mu_i(\bar{x} + \lambda(x^i-\bar{x})) \epsilon [C].$$

This implies that [C] is a cone with vertex $\bar{x}$, and by convexity of [C] the proof is complete.

We now focus our attention on cones that are related to some sets of interest. This leads to the following definitions and remarks.

Definition 20

Let A be a nonempty set in E^n and let $\bar{x} \in \bar{A}$. C_A is said to be a *cone of A at* $\bar{x}$ iff it is a cone with vertex $\bar{x}$ that contains A. C(A) is said to be the *minimal cone of A at* $\bar{x}$ iff it is a cone of A at $\bar{x}$ which is contained in all such cones.

The following remark constructs the minimal cone of a nonempty set A.

Remark 8

Let A be a nonempty set in E^n and let $\bar{x} \in \bar{A}$. Then $C(A) = \{\bar{x} + \lambda(x-\bar{x})\colon x \in A, \lambda \geq 0\}$.

Proof. It can be immediately checked that C(A) is a cone with vertex $\bar{x}$. By letting $\lambda = 1$ in the set identity, $C(A) = \{\bar{x} + \lambda(x-\bar{x})\colon x \in A, \lambda \geq 0\}$ it follows that $A \subset C(A)$. We need to show that if C_A is any cone of A at $\bar{x}$ then $C(A) \subset C_A$. Let $y \in C(A)$, then there exists an $x \in A$ and some $\lambda \geq 0$ such that $y = \bar{x} + \lambda(x-\bar{x})$. But since $A \subset C_A$, then $x \in C_A$, and since C_A is a cone with vertex $\bar{x}$, then $y = \bar{x} + \lambda(x-\bar{x}) \in C_A$. This implies that $C(A) \subset C_A$ and the proof is complete.

The following remark shows that the minimal cone of a convex set is indeed a convex cone. This motivates the definition of the minimal convex cone of a set at a boundary point, which follows the remark.

Remark 9

Let A be a nonempty convex set in E^n and $\bar{x} \in \bar{A}$. Then C(A) is a convex cone.

Proof. We let y^1 and $y^2 \in C(A)$ and then show that $y^1 + y^2 - \bar{x} \in C(A)$. By Remark 8, $y^1 = \bar{x} + \lambda_1(x^1-\bar{x})$ and $y^2 = \bar{x} + \lambda_2(x^2-\bar{x})$, where

$\lambda_1 \geq 0$, $\lambda_2 \geq 0$, x^1, and $x^2 \in A$. If either λ_1 or λ_2 is equal to zero, then the result is immediate. If on the other hand, λ_1 and λ_2 are positive, then let $\lambda_1 + \lambda_2 = \lambda$ and hence,

$$y^1 + y^2 - \bar{x} = \bar{x} + \lambda_1(x^1-\bar{x}) + \lambda_2(x^2-\bar{x}) = \bar{x} + \lambda((\lambda_1/\lambda)x^1 + (\lambda_2/\lambda)x^2 - \bar{x}).$$

By convexity of A the result follows immediately.

Definition 21

Let A be a nonempty set in E^n and $\bar{x}\epsilon\bar{A}$. A set is said to be a *convex cone of A at* $\bar{x}$ iff it is a convex cone with vertex $\bar{x}$ and contains A. A set is said to be the *minimal convex cone of A at* $\bar{x}$ iff it is a convex cone of A at $\bar{x}$ that is contained in all such cones.

In order to make use of existing theorems on convex cones, we focus our attention on the minimal convex cones of sets rather than the corresponding minimal cones. We first establish an important result that the minimal cone of the convex hull of a set is equivalent to the convex hull of the minimal cone of the set. Furthermore, we show that both cones are indeed equivalent to the minimal convex cone of the set. We then develop some results concerning dual cones of convex cones. This leads to a relationship between an outernormal to an intersection of sets and outernormals to the individual sets. This relationship constitutes the main result of this section and is used in Section 3.

Theorem 7

Let A be a nonempty set in E^n and $\bar{x}\epsilon\bar{A}$. Then $[C(A)] = C[A]$† and both are equal to the minimal convex cone of A at $\bar{x}$.

Proof. Let C be the minimal convex cone of A at $\bar{x}$. In order to prove the result we show that $C\subset C[A]\subset[C(A)]\subset C$. We first show that $C\subset C[A]$. Since $[A]$ is a convex set and $\bar{x}\epsilon\overline{[A]}$, then by Remark 9, $C[A] = \{\bar{x} + \lambda(x-\bar{x})\colon x\epsilon[A],\lambda \geq 0\}$ is a convex cone with vertex $\bar{x}$. But since $C[A]$ contains A, then $C\subset C[A]$. We now show that $C[A] \subset [C(A)]$. Let $y = \bar{x} + \lambda(x-\bar{x}) \epsilon C[A]$ where $\lambda \geq 0$ and $x\epsilon[A]$. Then $x = \sum_{i=1}^{n+1} \mu_i x^i$ where $\mu_i \geq 0$, $x^i\epsilon A$ $(i=1,2,\ldots,n+1)$ and $\sum_{i=1}^{n+1} \mu_i = 1$. Therefore, $y = \bar{x} + \lambda(x-\bar{x}) = \sum_{i=1}^{n+1} \mu_i(\bar{x} + \lambda(x^i-\bar{x}))$. But since $\bar{x} + \lambda(x^i-\bar{x}) \epsilon C(A)$ for each i, then $\sum_{i=1}^{n+1} \mu_i(\bar{x} + \lambda(x^i-\bar{x})) \epsilon [C(A)]$. This implies that $C[A] \subset [C(A)]$. Finally we show that $[C(A)] \subset C$. Let $y \epsilon [C(A)]$, then $y = \sum_{i=1}^{n+1} \mu_i x^i$ where $\mu_i \geq 0$, $x^i \epsilon C(A)$ for $i=1,2,\ldots,n+1$ and $\sum_{i=1}^{n+1} \mu_i = 1$. But since $C(A)\subset C$ then $x^i\epsilon C$ for $i=1,2,\ldots,n+1$, and by convexity of C it follows immediately that $y\epsilon C$. Therefore $[C(A)]\subset C$ and the proof is complete.

The following theorem gives the relationship between the minimal cone of an intersection of a finite number of convex sets and the minimal cones of the individual sets. The corollary to this theorem is used in Section 3 in developing the Fritz John necessary conditions.

---

†$C[A]$ is the minimal cone of the convex hull of A at $\bar{x}$. According to our notation the latter should be denoted by $C([A])$. To simplify notation, $C[A]$ is used, however.

Theorem 8

Let $B_o = \bigcap_{i=1}^{m} B_i$ be nonempty set in E^n and let $\bar{x}\epsilon\bar{B}_o$, where B_i $(i=1,2,\ldots,m)$ are convex sets in E^n. Then $C(B_o) = \bigcap_{i=1}^{m} C(B_i)$, where $C(B_i)$ is the minimal cone of B_i at $\bar{x}$ $(i=0,1,\ldots,m)$.

Proof. Since $B_o \subset B_i$, then it follows from Remark 8 that $C(B_o) \subset C(B_i)$ for $i=1,2,\ldots,m$. This further implies that $C(B_o) \subset \bigcap_{i=1}^{m} C(B_i)$. We now show that $\bigcap_{i=1}^{m} C(B_i) \subset C(B_o)$. Let $y \in \bigcap_{i=1}^{m} C(B_i)$, then $y = \bar{x} + \lambda_i(x^i-\bar{x})$ where $\lambda_i \geq 0$, $x^i \epsilon B_i$ for $i=1,2,\ldots,m$. If $\lambda_i = 0$ for some $i\epsilon\{1,2,\ldots,m\}$, then $y = \bar{x}$ and hence $y\epsilon C(B_o)$. If on the other hand $\lambda_i > 0$ for all $i\epsilon\{1,2,\ldots,m\}$, then let $\lambda_k = \max_i \lambda_i$. Therefore,

$$y = \bar{x} + \lambda_k(x^k-\bar{x}) = \bar{x} + \lambda_i(x^i-\bar{x}) \qquad \text{for } i=1,2,\ldots,m \tag{14}$$

Rearranging the terms in (14) it is immediate that,

$$x^k = (\lambda_i/\lambda_k)x^i + (1 - \lambda_i/\lambda_k)\bar{x} \qquad \text{for } i=1,2,\ldots,m \tag{15}$$

By convexity of B_i and since $(\lambda_i/\lambda_k) \epsilon (0,1]$ for all $i\epsilon\{1,2,\ldots,m\}$ it follows that $x^k \epsilon \bigcap_{i=1}^{m} B_i = B_o$. This implies that $y \epsilon C(B_o)$ and the proof is complete.

Corollary

Consider the nonempty sets $A_1,A_2,\ldots,A_m$ in E^n, and assume that $A = \bigcap_{i=1}^{m} A_i$ is nonempty. Let $\bar{x}\epsilon\bar{A}$ and assume that there exists a neighborhood N about $\bar{x}$ such that $[A\cap N] = \bigcap_{i=1}^{m} [A_i\cap N]$. Then

$C[A \cap N] = \bigcap_{i=1}^{m} C[A_i \cap N].$

The following remark gives a result about image cones. This result is used in developing the Fritz John conditions in Section 3.

Remark 10

Let A be a nonempty closed set in E^n and let $\bar{x} \in A$. Furthermore, suppose that $\bar{x} \in \partial[A^c \cap N]$ for some closed neighborhood N about $\bar{x}$. Then, $(\sim \overline{C[A^c \cap N]}) \cap N \subset A \cap N$.

Proof. Since $\bar{x} \in \partial[A^c \cap N]$ then $\bar{x} \in \partial\, C[A^c \cap N]$. Hence by convexity of $C[A^c \cap N]$ there exists a supporting hyperplane of $C[A^c \cap N]$ at $\bar{x}$. In other words there exists a nonzero vector $p \in E^n$ such that $\langle p, x-\bar{x}\rangle \leq 0$ for all $x \in \overline{C[A^c \cap N]}$. If $y \in (\sim \overline{C[A^c \cap N]}) \cap N$, then $y \in N$ and $y = \bar{x} + \lambda(x-\bar{x})$ for some $\lambda \leq 0$ where $x \in \overline{C[A^c \cap N]}$. Therefore $\langle p, y-\bar{x}\rangle = \lambda\langle p, x-\bar{x}\rangle \geq 0$. If $y \notin A \cap N$, then $y \notin A$. This implies that $y \in \operatorname{int}(A^c \cap N) \subset \operatorname{int} C[A^c \cap N]$ which further implies that $\langle p, x-\bar{x}\rangle < 0$, a contradiction. Therefore $y \in A \cap N$ and the proof is complete.

Under Definition 18 earlier we have defined a dual cone C^* of a cone C. We now present a definition of a dual cone of a set A. We then establish the equivalence of different dual cones related to a given set.

Definition 22

Let A be a nonempty set in E^n and $\bar{x} \in \bar{A}$. A set is said to be the *dual cone of A at $\bar{x}$* iff it is the set of all outernormals to A at $\bar{x}$.

Theorem 9

Let A be a nonempty set in E^n and $\bar{x} \in \bar{A}$. Then $C^* = C^*(A) = C^*[A]$, where C^* is the dual cone of A at $\bar{x}$, $C^*(A)$ is the dual cone of the minimal cone of A at $\bar{x}$, and $C^*[A]$ is the dual cone of the minimal convex cone of A at $\bar{x}$.

Proof. We show the result by proving that $C^* \subset C^*[A] \subset C^*(A) \subset C^*$. Since $C[A] \supset C(A) \supset A$, then by Remark 1, it follows immediately that $C^*[A] \subset C^*(A) \subset C^*$. Hence, it suffices to show that $C^* \subset C^*[A]$. Let $p \in E^n$ be an *o.n.* to A at $\bar{x}$, then $\langle p, x-\bar{x}\rangle \leq 0$ for all $x \in A$. Consider an arbitrary point $y \in C[A]$, then $y = \bar{x} + \lambda(x-\bar{x})$, where $\lambda \geq 0$ and $x \in [A]$. Therefore, $x = \sum_{i=1}^{n+1} \mu_i x^i$ where $\mu_i \geq 0$, $x^i \in A$ $(i=1,2,\ldots,n+1)$ and $\sum_{i=1}^{n+1} \mu_i = 1$. This implies that $\langle p, y-\bar{x}\rangle = \lambda \sum_{i=1}^{n+1} \mu_i \langle p, x^i-\bar{x}\rangle \leq 0$, and hence p is an *o.n.* to C[A] at $\bar{x}$. This completes the proof.

The following remark summarizes some results which are needed to prove the next theorem.

<u>Remark 11</u>

(i) Let C be a closed cone in E^n. Then $C^{**} = C$. If C is a cone, then C^* is a closed cone. See, for example [22].

(ii) If C_1 and C_2 are cones with a common vertex in E^n, then $(C_1+C_2)^* = C_1^* \cap C_2^*$, where $C_1 + C_2 = \{x\colon x = x^1 + x^2,\ x^1 \in C_1$ and $x^2 \in C_2\}$. See, for example [22].

(iii) Let C be a cone in E^n, then $(\bar{C})^* = C^*$. To show this recall that since $C \subset \bar{C}$, then $C^* \supset (\bar{C})^*$. On the other hand let $p \in E^n$ be an *o.n.* to C at $\bar{x}$ (the vertex of C). Consider an arbitrary $x \in \bar{C}$, then there exists a sequence $\{x^k\} \subset C$ such that $\lim_{k\to\infty} x^k = x$. Furthermore, since $\langle p, x^k-\bar{x}\rangle \leq 0$ for all k, then $\langle p, x-\bar{x}\rangle \leq 0$ and hence p is an *o.n.* to $\bar{C}$ at $\bar{x}$. This completes the proof.

The following theorem gives a necessary and sufficient condition that an *o.n.* to the intersection of two convex cones is the sum of the *o.n.*'s to the individual cones.

Theorem 10

Let C_1 and C_2 be convex cones in E^n with a common vertex. Then $(C_1 \cap C_2)^* = C_1^* + C_2^*$ iff $\overline{C_1 \cap C_2} = \bar{C}_1 \cap \bar{C}_2$.

Proof. We first assume that $(C_1 \cap C_2)^* = C_1^* + C_2^*$. By Remark 11, the following set of equalities can be concluded.

$$\bar{C}_1 \cap \bar{C}_2 = (\bar{C}_1)^{**} \cap (\bar{C}_2)^{**} = ((\bar{C}_1)^* + (\bar{C}_2)^*)^* = (C_1^* + C_2^*)^* = (C_1 \cap C_2)^{**}$$

$$= (\overline{C_1 \cap C_2})^{**} = \overline{C_1 \cap C_2}$$

Conversely we assume that $\overline{C_1 \cap C_2} = \bar{C}_1 \cap \bar{C}_2$. Then by Remark 11, we conclude the following equalities.

$$C_1^* + C_2^* = (\bar{C}_1)^* + (\bar{C}_2)^* = ((\bar{C}_1)^* + (\bar{C}_2)^*)^{**} = ((\bar{C}_1)^{**} \cap (\bar{C}_2)^{**})^*$$

$$= (\bar{C}_1 \cap \bar{C}_2)^* = (\overline{C_1 \cap C_2})^* = (C_1 \cap C_2)^*$$

This completes the proof.

Corollary

If C_i $(i=1,2,\ldots,m)$ are convex cones with a common vertex, then $\left(\bigcap_{i=1}^{m} C_i\right)^* = \sum_{i=1}^{m} C_i^*$ iff $\bar{C} = \bigcap_{i=1}^{m} \bar{C}_i$, where $C = \bigcap_{i=1}^{m} C_i$.

We now give a simplified sufficient condition for $\bigcap_{i=1}^{m} \bar{C}_i = \bar{C}$ to hold. If this new condition holds, then $\left(\bigcap_{i=1}^{m} C_i\right)^* = \sum_{i=1}^{m} C_i^*$.

Remark 12

Let C_i $(i=1,2,\ldots,m)$ be convex cones in E^n with a common vertex. If $\mathrm{int}(\bigcap_{i=1}^m C_i) \neq \phi$, then $\bar{C} = \bigcap_{i=1}^m \bar{C}_i$ where $C = \bigcap_{i=1}^m C_i$.

Proof. Since $C \subset C_i$, then $\bar{C} \subset \bar{C}_i$ for all $i \in \{1,2,\ldots,m\}$. This further implies that $\bar{C} \subset \bigcap_{i=1}^m \bar{C}_i$. We now show that $\bigcap_{i=1}^m \bar{C}_i \subset \bar{C}$. Let $x \in \bigcap_{i=1}^m \bar{C}_i$ and choose any $\bar{x} \in \mathrm{int}(\bigcap_{i=1}^m C_i)$. It follows immediately that $x(\lambda) = \lambda\bar{x} + (1-\lambda)x$ belongs to $\mathrm{int}(\bigcap_{i=1}^m C_i)$ for all $\lambda \in (0,1)$. But since $\bar{C}$ is a closed set, then $\lim_{\lambda \to o} x(\lambda) = x \in \bar{C}$. This completes the proof.

It should be noticed that the converse of the above remark is generally not true. This can be shown by the following example.

$$C_1 = \{(x,y)\colon x \in E^1,\ y \in E^1,\ y \geq |x|\}, \text{ and}$$

$$C_2 = \{(x,y)\colon x \in E^1,\ y \in E^1,\ y \leq -|x|\}.$$

It is immediate that $\overline{C_1 \cap C_2} = \bar{C}_1 \cap \bar{C}_2 = \{(0,0)\}$. On the other hand $\mathrm{int}(C_1 \cap C_2) = \phi$.

The following theorem gives a relationship between an *o.n.* to an intersection of a finite number of convex sets and outernormals to the sets.

Theorem 11

Let B_i $(i=1,2,\ldots,m)$ be nonempty convex sets in E^n such that $\mathrm{int}(B_o) \neq \phi$, where $B_o = \bigcap_{i=1}^m B_i$. Furthermore, let $\bar{x} \in \bar{B}_o$. Then $C^*(B_o) = \sum_{i=1}^m C^*(B_i)$, where $C^*(B_i)$ is the dual cone of the minimal cone of B_i at $\bar{x}$, for all $i \in \{0,1,2,\ldots,m\}$.

Proof. By Theorem 8, it follows that $C(B_o) = \bigcap_{i=1}^{m} C(B_i)$. Since $\text{int}(B_o) \neq \phi$, then $\text{int}(C(B_o)) \neq \phi$, and hence by Remark 12 and the corollary to Theorem 10, it follows that $C^*(B_o) = \sum_{i=1}^{m} C^*(B_i)$ and the proof is complete.

It should be noted that the above result may be stated as follows. $p^o \epsilon E^n$ is an *o.n.* to B_o at $\bar{x}$ iff $p^o = \sum_{i=1}^{m} p^i$ where p^i is an *o.n.* to B_i at $\bar{x}$ for all $i \epsilon \{1,2,\ldots,m\}$. It should also be noted that one may consider nonconvex sets by using the convex hulls of the sets under consideration. An example where similar results are developed for nonconvex sets is the development of the Fritz John conditions in Section 3.

3. Generalization of the Kuhn-Tucker and the Fritz John Conditions

The best known optimality criteria in nonlinear programming are the Fritz John and the Kuhn-Tucker conditions. Consider the nonlinear program, $\text{minimize}_x\{f_o(x)\colon f_i(x) \leq 0,\ i=1,2,\ldots,m\}$. The Fritz John conditions may be stated as follows: If $\bar{x}$ solves the above problem, then there exists an m+1 dimensional vector $u = (u_o,u_1,\ldots,u_m)$ such that,

$$f_i(\bar{x}) \leq 0,\ u_i f_i(\bar{x}) = 0 \qquad i=1,2,\ldots,m.$$

$$u \neq 0,\ u \geq 0$$

$$u_o \nabla f_o(\bar{x}) + \sum_{i=1}^{m} u_i \nabla f_i(\bar{x}) = 0$$

Kuhn and Tucker asserted that under some constraint qualification† at $\bar{x}$, u_o is positive. Further, if f_i $(i=0,1,\ldots,m)$ are convex and if the constraint set has a nonempty interior, then $u_o>0$ and the conditions are also sufficient for optimality. For a detailed discussion on the Fritz John and the Kuhn-Tucker conditions, one may refer to [35].

Different attempts to generalize the Kuhn-Tucker conditions have been undertaken. An extension of the Kuhn-Tucker conditions may be in the form of relaxing the differentiability assumption of the functions, weakening the constraint qualification, or enlarging the class of functions for which these conditions are also sufficient for optimality. Assuming that the objective function is convex, but not necessarily differentiable, and that the constraint functions are convex and Gateaux differentiable,†† and furthermore assuming that the constraint set has a nonempty interior, Rockafeller [43] gave an extension of the Kuhn-Tucker conditions in a locally convex Hausdorff topological vector space. This was accomplished via the conjugate

---

†Let I be the set of binding constraints at $\bar{x}$, i.e. $I = \{i: f_i(\bar{x}) = 0\}$. The constraint qualification is satisfied at $\bar{x}$ iff given a vector $t\epsilon E^n$ such that $\langle\nabla f_i(\bar{x}),t\rangle \leq 0$ for all $i\epsilon I$, there corresponds a differentiable arc $a(\mu)$ which is contained in the constraint set for all $\mu\epsilon[0,1]$. Moreover, $[da(\mu)/d\mu]_{\mu=0} = \lambda t$ where $\bar{x} = a(0)$ and $\lambda>0$. Different constraint qualifications may be found in [1], and [35].

††Let X be a vector space and Y be a normed space. Let $f: X\rightarrow Y$ be a mapping. If $\Delta f(x,u) = \lim_{\delta\rightarrow 0} 1/\delta\ [f(x+\delta u) - f(x)]$ exists for x and $u\epsilon X$ then $\Delta f(x,u)$ is called a Gateaux differential at x with increment u. If the limit exists for all $u\epsilon X$, f is said to be Gateaux differentiable at x. If f is Gateaux differentiable at all $x\epsilon X$, then f is said to be Gateaux differentiable.

function theory developed by Fenchel [18] and discussed in Chapter III. It is worthwhile mentioning that by confining the problem to E^n, the assumption of Gateaux differentiability is equivalent to the differentiability concept of Definition 9. By considering the directional derivatives of the objective function and the gradients of the constraint functions, Bram [7] presented a necessary condition for optimality. Different forms of the constraint qualification are discussed by Arrow, Hurwicz, and Uzawa [1]. Mangasarian [34] showed that the Kuhn-Tucker conditions are sufficient for a global minimum if the objective function in pseudo-convex and the constraint functions are quasi-convex. Finally, different generalizations of necessary optimality conditions in the case of both equality and inequality constraints were given by Canon, Cullum, and Polak [9] and by Mangasarian and Fromovitz [37].

Throughout this section the differentiability assumption of the objective function and the constraint functions is relaxed. Convexity is relaxed to supportability and continuity is also relaxed at some parts of the section. We first consider convex (locally convex) functions and develop a generalization of the Kuhn-Tucker conditions using the saddle value theorem. Subgradients play the role of the gradient vectors in the generalized optimality criteria, which are shown to be necessary and sufficient for a global (local) minimum. Secondly, we give an extension of the Fritz John necessary conditions if the functions are mildly qualified. The functions are assumed to be locally supportable either from below or from above and that all the functions are continuous in a neighborhood about the point under consideration. We then show that the above conditions reduce to the Kuhn-Tucker conditions if the constraint functions and the objective function are supportable from below and if the constraint set has a nonempty interior. We also show that these conditions are also sufficient. Hence we obtain a further generalization of the results obtained via the saddle value theorem. As a corollary to a more general result, we then give a necessary and sufficient condition for optimality when the point under consideration is an interior feasible solution. The constraint functions are assumed to be continuous but no restrictions are imposed on the objective function. We also consider two problems, an equality constrained problem and an inequality constrained problem, which are both equivalent to the original problem. This leads to the last theorems of this chapter which give sufficient conditions for local optimality with only the objective function constrained to be locally supportable from below. The conditions in Theorem 16 require that the lagrangian multiplier associated with the objective function should be positive, and thus the results form an extension of the sufficiency of the Kuhn-Tucker conditions. We also show that if the objective function is pseudo-convex and the constraints are quasi-convex, then the same conditions assure global optimality. This is a generalization of a result by Mangasarian [34] for differentiable functions. It may be noted that the sufficient conditions in Theorem 17 may be used even if the lagrangian multiplier associated with the objective function is zero. In Appendix B we give some examples of the application of this theorem. The familiar problem of the outward cusp which does not satisfy the Kuhn-Tucker constraint qualification is considered and we show that the point under consideration satisfies one of our sufficient conditions and hence is optimal.

Generalization of the Kuhn-Tucker Conditions for Convex Functions

The problem under consideration is to minimize$_x\{f_o(x)$: $f_i(x) \leq 0$, $i=1,2,\ldots,m\}$, where $f_o,f_1,\ldots,f_m$: $E^n \to E^1$ are convex functions. By making use of the saddle value Theorem [5], and the corollary to Theorem 6, we give an extension of the Kuhn-Tucker conditions by relaxing differentiability.

Theorem 12

Let f_i: $E^n \to E^1$ $(i=0,1,2,\ldots,m)$ be convex functions and assume that there exists an $x \epsilon E^n$ such that $f_i(x) < 0$ for all $i \epsilon \{1,2,\ldots,m\}$. If $\bar{x} \epsilon E^n$ is a feasible solution of the problem P: minimize$_x\{f_o(x)$: $f_i(x) \leq 0$, $i=1,2,\ldots,m\}$ then $\bar{x}$ is an optimal† solution of P iff there exists a vector $\bar{u} = (\bar{u}_1,\bar{u}_2,\ldots,\bar{u}_m)$ such that,

$$\bar{u}_i \geq 0, \ \bar{u}_i f_i(\bar{x}) = 0 \qquad i=1,2,\ldots,m$$

$$\theta^o(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \theta^i(\bar{x}) = 0$$

where $\theta^i(\bar{x})$ is a subgradient of f_i at $\bar{x}$, for all $i \epsilon \{0,1,\ldots,m\}$.

---

†$\bar{x}$ is an optimal solution of problem P iff $f_i(\bar{x}) \leq 0$ for all $i \epsilon \{1,2,\ldots,m\}$ and $f_o(\bar{x}) \leq f_o(x)$ for all x such that $f_i(x) \leq 0$ for all $i \epsilon \{1,2,\ldots,m\}$.

Proof. We first show that the conditions are sufficient. The following set of inequalities is obtained from Theorem 4.

$$f_o(x) \geq f_o(\bar{x}) + \langle\theta^o(\bar{x}), x-\bar{x}\rangle \qquad \text{for all } x\epsilon E^n \tag{16}$$

$$f_i(x) \geq f_i(\bar{x}) + \langle\theta^i(\bar{x}), x-\bar{x}\rangle \qquad \text{for all } x\epsilon E^n, \text{ all } i\epsilon\{1,2,\ldots,m\} \tag{17}$$

Multiplying (17) by $\bar{u}_i$ and adding (16) and the m inequalities corresponding to (17) we conclude that,

$$\begin{aligned} f_o(x) + \sum_{i=1}^{m} \bar{u}_i f_i(x) &\geq f_o(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i f_i(\bar{x}) \\ &+ \langle\theta^o(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \theta^i(\bar{x}), x-\bar{x}\rangle = f_o(\bar{x}) \end{aligned} \tag{18}$$

If $x\epsilon E^n$ is a feasible solution of problem P, then $f_i(x) \leq 0$ for all $i\epsilon\{1,2,\ldots,m\}$. But since $\bar{u} \geq 0$, then $\sum_{i=1}^{m} \bar{u}_i f_i(x) \leq 0$. Therefore (18) implies that $f_o(x) \geq f_o(\bar{x})$ for all feasible x, and hence $\bar{x}$ is an optimal solution of problem P. To show that the conditions are necessary, assume that $\bar{x}$ solves problem P and consider $\psi(x,u) = f_o(x) + \langle u, f(x)\rangle$, where $f(x) = (f_1(x), f_2(x), \ldots, f_m(x))$. Given that there exists an $x\epsilon E^n$ such that $f(x) < 0$, it can be shown that $\bar{x}$ solves problem P iff there exists $\bar{u} \geq 0$ such that,

$$\psi(\bar{x},u) \leq \psi(\bar{x},\bar{u}) \leq \psi(x,\bar{u}) \qquad \text{for all } x\epsilon E^n,\ 0\leq u\epsilon E^m \tag{19}$$

See Berge and Ghoulia-Houri [5]. We first show that $\bar{u}_i f_i(\bar{x}) = 0$ for all $i \in \{1,2,\ldots,m\}$. Assume on the contrary that $\bar{u}_j f_j(\bar{x}) < 0$ for some j, then $\bar{u}_j > 0$ and $f_j(\bar{x}) < 0$. Consider $(\bar{x},u) \in E^{n+m}$, where $u_i = \bar{u}_i$ for $i \neq j$ and $u_j = \bar{u}_j/2$. Therefore we conclude that,

$$\psi(\bar{x},u) = f_o(\bar{x}) + \langle u,f(\bar{x})\rangle = f_o(\bar{x}) + \sum_{\substack{i=1\\ i\neq j}}^{m} \bar{u}_i f_i(\bar{x}) + (\bar{u}_j/2)f_j(\bar{x})$$

$$= f_o(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i f_i(\bar{x}) - (\bar{u}_j/2)f_j(\bar{x})$$

$$> f_o(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i f_i(\bar{x}) = \psi(\bar{x},\bar{u}).$$

This contradicts (19) and hence $\bar{u}_i f_i(\bar{x}) = 0$ for all $i \in \{1,2,\ldots,m\}$. We now consider the set $A = \{(x,y)\colon x \in E^n,\ y \in E^1,\ y \geq \psi(x,\bar{u})\}$. From (19) it follows that $(\bar{x},\psi(\bar{x},\bar{u}))$ is a minimum point of A, and hence by Remark 3, $(0,-1) \in E^{n+1}$ is an *o.n.* to A at $(\bar{x},\psi(\bar{x},\bar{u}))$. This implies that $0 \in E^n$ is a subgradient of $\psi(.,\bar{u})$ at $\bar{x}$. But since $\bar{u} \geq 0$ and $f_i (i=0,1,\ldots,m)$ are convex functions, then $\psi(.,\bar{u})$ is a convex function. Therefore by the corollary to Theorem 6, we conclude that $\theta^o(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \theta^i(\bar{x}) = 0$, where $\theta^i(\bar{x})$ is a subgradient of f_i at $\bar{x}$ for all $i \in \{0,1,2,\ldots,m\}$. This completes the proof.

It should be noted that if the functions $f_o, f_1, \ldots,$ and f_m are locally convex at $\bar{x}$, then the above conditions are necessary and sufficient for a local minimum. Furthermore, if the functions are differentiable, then by Theorem 3 the subgradients are reduced to the gradient vectors and we obtain the Kuhn-Tucker conditions as a special case.

Generalization of the Fritz John and the Kuhn-Tucker Conditions to Supportable Functions

The problem under consideration is to minimize$_x$$\{f_o(x)$: $f_i(x) \leq 0$, $i=1,2,\ldots,m\}$. All the functions are assumed to be continuous. Under some mild qualification we develop a generalization of the Fritz John necessary conditions for optimality. This result is given by Theorem 13 below, where the qualification is given by the existence of a neighborhood N, about the point $\bar{x}$ under investigation, that satisfies conditions (i), (ii), and (iii). It may be helpful to briefly discuss these conditions. Condition (i) states that the objective function and each of the binding constraints is either locally supportable from below or from above at the point under consideration. Condition (ii) is automatically satisfied for sufficiently small neighborhoods since the constraint functions are continuous at the point under investigation. It should be noticed that if condition (i) and (ii) are satisfied by a neighborhood N, then they are also satisfied by any $\bar{N}\subset N$. Note that for all $i\epsilon J$ the functions f_i are locally supportable from below at $\bar{x}$ and for all $i\epsilon K$ the functions f_i are locally supportable from above at $\bar{x}$ and hence $J\cup K = I\cup\{0\}$. In case some function f_j is locally supportable from both above and below at $\bar{x}$, then we want j to be included in only one set J or K. Hence the assumption $J\cap K = \phi$ is required in the theorem below. By considering the image cones of $[B_i\cap N]$ for all $i\epsilon K$, where B_i is the hypograph of f_i and N is some neighborhood about $(\bar{x},o)$, we are assured by Remark 10 that $(\sim C[B_i\cap N])\cap N\subset A_i\cap N$ where A_i is the epigraph of f_i. This motivates the introduction of the set A in Theorem 13 below, which is locally supportable at $(\bar{x},o)$ by construction. Condition (iii) below asserts that $[A \cap N]$ can be expressed as the intersection of convex sets and hence Theorem 11 can be applied. There is some reason to believe that this condition is satisfied by choosing N sufficiently small, at least for "well behaving" functions.

Theorem 13

Consider the functions $f_o, f_1, \ldots, f_m$: $E^n \to E^1$ and assume that f_i $(i=0,1,2,\ldots,m)$ are continuous. Consider the problem P: $\text{minimize}_x\{f_o(x)\colon f_i(x) \le 0,\ i=1,2,\ldots,m\}$ and let $\bar{x} \in E^n$ be a feasible solution of problem P. Let $I = \{i\colon f_i(\bar{x})=0\}$,

$$A_i = \{(x,y)\colon x \in E^n,\ y \in E^1,\ y \ge f_i(x) - f_i(\bar{x})\},\ \text{and}$$

$$B_i = \{(x,y)\colon x \in E^n,\ y \in E^1,\ y \le f_i(x) - f_i(\bar{x})\}$$

for all $i \in I \cup \{0\}$. Suppose that there exists a closed neighborhood N about $(\bar{x},o)$ such that,

(i) $(\bar{x},o) \in (\partial[A_i \cap N]) \cup (\partial[B_i \cap N])$ for all $i \in I \cup \{0\}$.

(ii) $f_i(x) \le 0$ for all x such that $(x,y) \in N$, for all $i \notin I$ and $i \ne o$.

(iii) $[A \cap N] = (\bigcap_{i \in J} [A_i \cap N]) \cap (\bigcap_{i \in K} (\sim C[B_i \cap N]) \cap N)$, where

$$A = (\bigcap_{i \in J} A_i) \cap (\bigcap_{i \in K} (\sim C[B_i \cap N])).$$

$$J = \{i\colon (\bar{x},o) \in \partial[A_i \cap N]\} \subset I \cup \{0\}$$

$$K = \{i\colon (\bar{x},o) \in \partial[B_i \cap N]\} \subset I \cup \{0\}$$

$$J \cup K = I \cup \{0\} \quad \text{and} \quad J \cap K = \phi.$$

If $\bar{x}$ is an optimal solution of problem P, then there exists a nonzero vector $\bar{u} = (\bar{u}_o, \bar{u}_1, \ldots, \bar{u}_m)$ such that,

$$\bar{u}_o \geq 0, \quad \bar{u}_i \geq 0 \quad \bar{u}_i f_i(\bar{x}) = 0 \qquad \text{for } i=1,2,\ldots,m.$$

$$\bar{u}_o \theta^o(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \theta^i(\bar{x}) = 0$$

where $\theta^i(\bar{x})$ is a subgradient of f_i at $\bar{x}$ for $i=0,1,2,\ldots,m$.

Proof. Let $(x,y) \in A \cap N$. If x is a feasible solution of problem P, then $f_o(x) \geq f_o(\bar{x})$ since $\bar{x}$ is an optimal solution of this problem. If f_o is locally supportable from below at $\bar{x}$, then $o \in J$ and hence $(x,y) \in A_o$. Therefore by definition, $y \geq f_o(x) - f_o(\bar{x}) \geq 0$. If f_o is locally supportable from above at $\bar{x}$, then $(x,y) \in {\sim}C[B_o \cap N] \cap N$, and hence by Remark 10, $(x,y) \in A_o \cap N$, and again $y \geq 0$. If on the other hand x is not feasible, then by assumption (ii) it follows that $f_i(x) > 0$ for some $i \in I$. A similar argument shows that $y > o$. At any rate we showed that if $(x,y) \in A \cap N$, then $y \geq o$. This implies that $(\bar{x},o)$ is a minimum point of $A \cap N$, and hence by Remark 3 $(0,-1) \in E^{n+1}$ is an *o.n.* to $A \cap N$ at $(\bar{x},o)$. By Theorem 9 it follows that $(0,-1)$ is an *o.n.* to $[A \cap N] = (\bigcap_{i \in J} [A_i \cap N]) \cap (\bigcap_{i \in K} ({\sim}C[B_i \cap N]) \cap N)$ at $(\bar{x},o)$. It can be shown that $(\bar{x},\delta) \in \operatorname{int}([A_i \cap N])$ for all $i \in J$ and that $(\bar{x},\delta) \in \operatorname{int}(({\sim}C[B_i \cap N]) \cap N)$ for all $i \in K$, when $\delta > o$ is sufficiently small. Therefore, $\operatorname{int}([A \cap N]) \neq \phi$ and hence all the hypotheses of Theorem 11 are satisfied. Therefore,

$$(0,-1) = \sum_{i\epsilon J} p^i + \sum_{i\epsilon K} p^i \tag{20}$$

where p^i is an *o.n.* to $[A_i \cap N]$ at $(\bar{x},o)$ for all $i\epsilon J$ and p^i is an *o.n.* to $(\sim C[B_i \cap N]) \cap N$ at $(\bar{x},o)$ for all $i\epsilon K$. From Equation (20) it is immediate that $p^i \neq 0$ for some $i\epsilon J \cup K$. We therefore let $J_o = \{i\colon i\epsilon J, p^i \neq 0\}$ and $K_o = \{i\colon i\epsilon K, p^i \neq 0\}$. Therefore, by Theorem 9 it follows that p^i is an *o.n.* to $A_i \cap N$ at $(\bar{x},o)$ for $i\epsilon J_o$, and by Definition 17 and Theorem 9, it follows that $-p^i$ is an *o.n.* to $B_i \cap N$ at $(\bar{x},o)$ for $i\epsilon K_o$. Let $p^i = (\bar{p}^i, p^i_{n+1})$ for all $i\epsilon J_o \cup K_o$. We assert that $p^i_{n+1} < o$ for all $i\epsilon J_o \cup K_o$. We show that p^i_{n+1} is neither zero nor positive if $i\epsilon J_o \cup K_o$. Assume on the contrary that $p^i_{n+1} = 0$ for some $i\epsilon J_o$. Then $p^i_j \neq 0$ for some $j\epsilon\{1,2,\ldots,n\}$. Consider the point $(x,f_i(x))\epsilon A_i \cap N$ where $x_k = \bar{x}_k$ for all $k \neq j$ and $x_j = \bar{x}_j + \delta p^i_j$, where $\delta > o$ is sufficiently small such that $(x,f_i(x))\epsilon A_i \cap N$. Therefore, $\langle p^i,(x,f_i(x))-(\bar{x},o)\rangle = \langle \bar{p}^i, x-\bar{x}\rangle = \delta(p^i_j)^2 > 0$ which contradicts the fact that p^i is an *o.n.* to $A_i \cap N$ at $(\bar{x},o)$. Therefore $p^i_{n+1} \neq 0$ for $i\epsilon J_o$. We show that p^i_{n+1} is not positive when $i\epsilon J_o$. Assume on the contrary that $p^i_{n+1} > 0$, then consider $(\bar{x},\delta)\epsilon A_i \cap N$ where $\delta > 0$ is sufficiently small to insure that $(\bar{x},\delta)\epsilon N$. Therefore,

$$\langle p^i,(\bar{x},\delta) - (\bar{x},o)\rangle = p^i_{n+1}\delta > 0$$

which is again a contradiction. This implies that $p^i_{n+1} < 0$ for all $i\epsilon J_o$. A similar argument shows that this is also the case for all $i\epsilon K_o$. Hence we can write Equation (20) as follows.

$$(0,-1) = \sum_{i\in J\cup K} p^i = \sum_{i\in J_o\cup K_o} p^i = \sum_{i\in J_o\cup K_o} (\bar{p}^i, p^i_{n+1})$$

$$= \sum_{i\in J_o\cup K_o} |p^i_{n+1}|(\theta^i(\bar{x}),-1)$$

$$= \sum_{i\in J_o\cup K_o} u_i(\theta^i(\bar{x}),-1) \tag{21}$$

where $u_i = |p^i_{n+1}| > 0$, and $\theta^i(\bar{x}) = \bar{p}^i/|p^i_{n+1}|$ for all $i\epsilon J_o\cup K_o$. It follows immediately from Remark 1 that $(\theta^i(\bar{x}),-1)$ is an *o.n.* to $A_i\cap N$ at $(\bar{x},o)$ and hence $\theta^i(\bar{x})$ is a subgradient of f_i at $\bar{x}$ for all $i\epsilon J_o$. Also $(\theta^i(\bar{x}),-1)$ is an *o.n.* to $(\sim C[B_i\cap N])\cap N$ at $(\bar{x},o)$ and hence $(-\theta^i(\bar{x}),1)$ is an *o.n.* to $B_i\cap N$ at $(\bar{x},o)$ which further implies that $\theta^i(\bar{x})$ is a subgradient of f_i at $\bar{x}$ for all $i\epsilon K_o$. Considering the first n components of the vector Equation (21), we obtain the following.

$$0 = \sum_{i\epsilon J_o\cup K_o} u_i\theta^i(\bar{x}) = \sum_{i=0}^{m} u_i\theta^i(\bar{x}) = u_o\theta^o(\bar{x}) + \sum_{i=1}^{m} u_i\theta^i(\bar{x})$$

where we let $u_i = o$ for $i\notin J_o\cup K_o$. It follows immediately that $u = (u_o,u_1,\ldots,u_m) \neq o$, and $u_i \geq o$ for all $i\epsilon\{0,1,2,\ldots,m\}$. Since $f_i(\bar{x}) = 0$ for all $i\epsilon J_o\cup K_o$, then $u_i f_i(\bar{x}) = 0$ for all $i\epsilon\{1,2,\ldots,m\}$. This completes the proof.

We now consider the question, under what conditions is u_o positive? Recall that if $u_o>0$, then we can divide the vector equation $u_o\theta^o(\bar{x}) + \sum_{i=1}^{m} u_i\theta^i(\bar{x}) = 0$ by u_o and obtain the Kuhn-Tucker conditions. This is dealt with in Theorem 14, the proof of which makes use of the following result.

Remark 13

Consider the functions $f_1, f_2, \ldots, f_m$: $E^n \to E^1$. Let $\bar{x} \in E^n$ and $I = \{i : f_i(\bar{x}) = 0\}$. Suppose that there exists some $x^o \in E^n$ such that $f_i(x^o) < 0$ and that f_i is supportable from below at $\bar{x}$ for all $i \in I$. Let $\theta^i(\bar{x})$ be a subgradient of f_i at $\bar{x}$ for $i=1,2,\ldots,m$. Then the system $\sum_{i=1}^{m} u_i \theta^i(\bar{x}) = 0$ has no nonzero solution $u = (u_1, u_2, \ldots, u_m)$ such that $u_i \geq 0$ and $u_i f_i(\bar{x}) = 0$ for all $i \in \{1,2,\ldots,m\}$.

Proof. Assume on the contrary that such a solution exists. Then $u_i = 0$ for all $i \notin I$. Since f_i is supportable from below at $\bar{x}$, then by Theorem 1 we conclude that

$$f_i(x) \geq f_i(\bar{x}) + \langle \theta^i(\bar{x}), x-\bar{x} \rangle = \langle \theta^i(\bar{x}), x-\bar{x} \rangle \tag{22}$$

$$\text{for all } x \in E^n, \text{ all } i \in I.$$

By letting $x = x^o$ in (22) it follows that $\langle \theta^i(\bar{x}), x^o-\bar{x} \rangle < 0$ for all $i \in I$. Furthermore since $u \neq o$, then for some $i \in I$, $u_i > 0$. Therefore it follows that $\sum_{i=1}^{m} u_i \langle \theta^i(\bar{x}), x^o-\bar{x} \rangle < 0$ which contradicts that $\sum_{i=1}^{m} u_i \theta^i(\bar{x}) = 0$. This completes the proof.

Theorem 14 below gives a further generalization of the Kuhn-Tucker conditions. We show, by the aid of the above remark, that in Theorem 13 if all the binding constraints are supportable from below,[†] and if there exists on x^o such that $f_i(x^o) < 0$ for all $i \in I$, then $u_o > 0$.

[†]Note that with this assumption the set $K = \phi$ in Theorem 13.

Furthermore we show that the conditions are also sufficient for optimality.

Theorem 14

Consider the functions $f_o, f_1, \ldots, f_m$: $E^n \to E^1$ and assume that f_i $(i=0,1,2,\ldots,m)$ are continuous. Consider the problem P: $\text{minimize}_x \{f_o(x): f_i(x) \le 0,\ i=1,2,\ldots,m\}$ and let $\bar{x} \in E^n$ be a feasible solution of problem P. Let $I = \{i: f_i(\bar{x}) = 0\}$. Furthermore let $A_i = \{(x,y): x \in E^n,\ y \in E^1,\ y \ge f_i(x) - f_i(\bar{x})\}$ for all $i \in I \cup \{0\}$. Suppose that there exists a closed neighborhood N about $(\bar{x}, o)$ such that,

(i) $(\bar{x}, o) \in \partial[A_i \cap N]$ for all $i \in I \cup \{0\}$.

(ii) $f_i(x) \le 0$ for all x such that $(x,y) \in N$, for all $i \notin I$ and $i \ne 0$.

(iii) $[A \cap N] = \bigcap_{i \in I \cup \{0\}} [A_i \cap N]$ where $A = \bigcap_{i \in I \cup \{0\}} A_i$.

Assume that there exists an $(x^o, y) \in N$ such that $f_i(x^o) < 0$ for all $i \in I$. Then $\bar{x}$ is a local minimum solution to problem P above iff there exists a vector $\bar{u} = (\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m)$ such that,

$$\bar{u}_i \ge 0 \qquad \bar{u}_i f_i(\bar{x}) = 0 \qquad \text{for } i=1,2,\ldots,m.$$

$$\theta^o(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \theta^i(\bar{x}) = 0,$$

where $\theta^i(\bar{x})$ is a subgradient of f_i at $\bar{x}$, for all $i \in I \cup \{0\}$.

Proof. The hypotheses of Theorem 13 are satisfied and hence there exists a vector $(u_o, u_1, \ldots, u_m) \ne 0$ such that,

$$u_o \ge 0 \qquad u_i \ge 0 \qquad u_i f_i(\bar{x}) = 0 \qquad \text{for } i=1,2,\ldots,m$$

$$u_o \theta^o(\bar{x}) + \sum_{i=1}^{m} u_i \theta^i(\bar{x}) = 0$$

We assert that $u_o > 0$, because if $u_o = 0$ an immediate contradiction of Remark 13 follows. So if we divide the above vector equation by u_o and denote u_i/u_o by $\bar{u}_i$ the necessary part follows. To show the sufficiency part let $\bar{N} = \{x: (x,y) \in A \cap N\}$. Recall that condition (i) above implies supportability from below and hence by Theorem 1, it follows that,

$$f_o(x) \geq f_o(\bar{x}) + \langle \theta^o(\bar{x}), x-\bar{x} \rangle \qquad \text{for all } x \in \bar{N} \tag{23}$$

$$f_i(x) \geq f_i(\bar{x}) + \langle \theta^i(\bar{x}), x-\bar{x} \rangle \quad \text{for all } x \in \bar{N}, \text{ all } i \in I \tag{24}$$

Multiplying (24) by $\bar{u}_i \geq 0$ and adding (23) and the inequalities corresponding to (24) we obtain the following.

$$\begin{aligned} f_o(x) + \sum_{i \in I} \bar{u}_i f_i(x) &\geq f_o(\bar{x}) + \sum_{i \in I} f_i(\bar{x}) + \langle (\theta^o(\bar{x}) + \sum_{i \in I} \bar{u}_i \theta^i(\bar{x})), x-\bar{x} \rangle \\ &= f_o(\bar{x}) \qquad\qquad \text{for all } x \in \bar{N} \end{aligned} \tag{25}$$

Therefore if x is a feasible solution, then $f_i(x) \leq 0$, and hence $\sum_{i \in I} \bar{u}_i f_i(x) \leq 0$. The result follows immediately then from (25).

The following theorem gives a necessary and sufficient condition for optimality when the point $(\bar{x}, o) \in E^{n+1}$ is an interior point of the epigraphs of $f_1, f_2, \ldots,$ and f_m. No continuity assumption of the functions involved is required. As a corollary to this theorem, we give a necessary and sufficient condition for optimality when $\bar{x}$ is an interior point of the feasible region and the constraint functions are continuous. It should be noted that it follows immediately from Theorem 14 above that a necessary and sufficient condition for optimality in such a case is that $\theta^{o}(\bar{x}) = 0$. In the corollary that follows Theorem 15 below, no supportability of the objective function is assumed, however.

Theorem 15

Consider the problem P: minimize$_x\{f_o(x): f_i(x) \le 0,\ i=1,2,\ldots,m\}$ where $f_i: E^n \to E^1$, $i=0,1,2,\ldots,m$. Let $(\bar{x},o)\epsilon E^{n+1}$ be an interior point of the epigraph of f_i, for each $i\epsilon\{1,2,\ldots,m\}$. A necessary and sufficient condition that $\bar{x}$ is a local (global) minimum solution to problem P, is that $(0,-1)\epsilon E^{n+1}$ is a *l.o.n.* *(o.n.)* to the set $A_o = \{(x,y): x\epsilon E^n, y\epsilon E^1, y \ge f_o(x)-f_o(\bar{x})\}$ at $(\bar{x},0)$.

Proof. The sufficiency part is trivial. We show that the condition is necessary by showing that if $(0,-1)$ is not a *l.o.n.* to A_o at $(\bar{x},0)$, then $\bar{x}$ does not solve the problem locally. We first assert that there exists a neighborhood N about $\bar{x}$ such that $f_i(x) < 0$ for all $x\epsilon N$ and all $i\epsilon\{1,2,\ldots,m\}$. Such a neighborhood is constructed as follows. Since $(\bar{x},0)\epsilon \text{int}(A_i)$, where A_i is the epigraph of f_i for each $i\epsilon\{1,2,\ldots,m\}$, then there exists a neighborhood $\bar{N}$ about $(\bar{x},0)$ such that,

$$\bar{N} = \{(x,y): x\epsilon E^n, y\epsilon E^1, \|(x,y) - (\bar{x},0)\| < \epsilon\} \subset \bigcap_{i=1}^{m} \text{int}(A_i)$$

where $\epsilon > 0$. Consider the neighborhood N about $\bar{x}$ given by,

$N = \{x\colon x \in E^n,\ \|x-\bar{x}\| < \epsilon/2\}$. If $x \in N$, then by triangle inequality it follows that,

$$\|(x,-\epsilon/2) - (\bar{x},o)\| \leq \|(x,o) - (\bar{x},o)\|$$

$$+ \|(x,-\epsilon/2) - (x,o)\| < \epsilon/2 + \epsilon/2 = \epsilon.$$

This implies that $(x,-\epsilon/2)\in\bar{N}$, and hence $-\epsilon/2 \geq f_i(x)$ for each $i\in\{1,2,\ldots,m\}$. This proves the above assertion. Since $(0,-1)$ is assumed not to be a *l.o.n.* to A_o at $(\bar{x},0)$, then for each neighborhood M about $(\bar{x},0)$ there exists an x (which depends upon M) such that $(x,f_o(x) - f_o(\bar{x}))\in M$ and $f_o(x) < f_o(\bar{x})$. By choosing an M of radius $\epsilon/2$, a contradiction is immediate. This completes the proof.

Corollary

Consider the problem P: minimize$_x\{f_o(x)\colon f_i(x) \leq 0,\ i=1,2,\ldots,m\}$. Let $f_1,f_2,\ldots,$ and f_m be continuous, and suppose that $f_i(\bar{x}) < 0$ for each $i\in\{1,2,\ldots,m\}$. Then $\bar{x}$ is a local (global) minimum solution to problem P iff $(0,-1)$ is a *l.o.n. (o.n.)* to $A_o = \{(x,y)\colon x\in E^n,\ y\in E^1,\ y\geq f_o(x)-f_o(\bar{x})\}$ at $(\bar{x},0)$.

Proof. Proof follows immediately from the above theorem, by noting that under the mentioned hypotheses $(\bar{x},0)$ is an interior point of the epigraph of f_i for each $i\in\{1,2,\ldots,m\}$.

It should be noticed that the continuity of the constraints is essential to prove the necessary part of the above corollary. We give the following counter example when continuity is relaxed. Consider the problem, $\text{minimize}_x\{-x\colon f(x)\le 0\}$ where

$$f(x) = \begin{cases} x - 1 & \text{if } x\le 0 \\ x + 1 & \text{if } x>0 \end{cases}$$

clearly the optimal point is $x = 0$. However, $(0,-1)$ is not a *l.o.n.* to the epigraph of $-x$ at $(0,0)$.

Further Optimality Conditions

It may be recalled that the optimality conditions derived earlier assume that the binding constraints and the objective function are supportable either from above or below. Here we relax the supportability assumption of the constraints. The resulting conditions for optimality are different from the earlier results.

The optimality conditions to be developed are obtained by considering an inequality constrained problem which is equivalent to the problem under consideration. The definition of the new constraint functions insures the existence of supporting hyperplanes. We show that every feasible solution satisfies the necessary conditions and hence they are useless. We focus our attention on sufficient conditions.

The following remark gives the forms of an equality constrained problem and an inequality constrained problem which are both equivalent to the original problem.

Remark 14

The following three problems are equivalent.

(i) $\text{minimize}_x\{f_o(x): f_i(x) \leq 0, i=1,2,\ldots,m\}$

(ii) $\text{minimize}_x\{f_o(x): g_i(x) = 0, i=1,2,\ldots,m\}$.

(iii) $\text{minimize}_x\{f_o(x): g_i(x) \leq 0, i=1,2,\ldots,m\}$

where $g_i = \max(0,f_i)$ for all $i\epsilon\{1,2,\ldots,m\}$.

The result is immediate by noting that $f_i(x) \leq 0$ implies that $g_i(x)= 0$ which implies that $g_i(x) \leq 0$, which in turn implies that $f_i(x) \leq 0$, for all $x\epsilon E^n$, all $i\epsilon\{1,2,\ldots,m\}$.

We now show that the Fritz John conditions corresponding to the problem, $\text{minimize}_x\{f_o(x): g_i(x) \leq 0, i=1,2,\ldots,m\}$ are useless since every feasible solution satisfies them.

Remark 15

Consider the functions $f_o,f_1,\ldots,f_m$: $E^n \to E^1$ and let $g_i = \max(0,f_i)$ for all $i\epsilon\{1,2,\ldots,m\}$. Consider the problem, $\text{minimize}_x\{f_o(x): f_i(x) \leq 0, i=1,2,\ldots,m\}$ and the equivalent problem, $\text{minimize}_x\{f_o(x): g_i(x) \leq 0, i=1,2,\ldots,m\}$. Consider an arbitrary feasible solution $\bar{x}\epsilon E^n$. Let $A_o = \{(x,y): x\epsilon E^n, y\epsilon E^1, y \geq f_o(x) - f_o(\bar{x})\}$ and $A_i = \{(x,y): x\epsilon E^n, y\epsilon E^1, y \geq g_i(x)\}$ for all $i\epsilon\{1,2,\ldots,m\}$. Then there exists a nonzero vector $\bar{u} = (\bar{u}_o,\bar{u}_1,\ldots,\bar{u}_m)$ such that,

$$\bar{u}_o \geq 0 \qquad \bar{u}_i \geq 0 \qquad \bar{u}_i g_i(\bar{x}) = 0 \qquad \text{for } i=1,2,\ldots,m.$$

$$\bar{u}_o \theta^o(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \theta^i(\bar{x}) = 0$$

where $\theta^i(\bar{x})$ is a subgradient of g_i at $\bar{x}$ for all $i \in \{1,2,\ldots,m\}$ and $\theta^o(\bar{x})$ is a subgradient of f_o at $\bar{x}$.

Proof. Since $\bar{x}$ is a feasible solution, then $f_i(\bar{x}) \leq 0$ and hence $g_i(\bar{x}) = 0$ for all $i \in \{1,2,\ldots,m\}$. For all $x \in E^n$, $g_i(x) \geq 0$ and hence it follows that $(\bar{x},0)$ is a minimum point of A_i, for all $i \in \{1,2,\ldots,m\}$. This further implies that $(0,-1)$ is an *o.n.* to A_i at $(\bar{x},0)$ and hence $\theta^i(\bar{x}) = 0$ for all $i \in \{1,2,\ldots,m\}$. Therefore if we let $\bar{u}_o = 0$ and $\bar{u}_i = 1$ for all $i \in \{1,2,\ldots,m\}$, the above conditions are immediately satisfied.

We now consider sufficient optimality conditions. The sufficient conditions of Theorem 16 below can be applied for any feasible solution. The supportability from below of the objective function is needed but no such assumption is required with regard to the constraint functions.

<u>Theorem 16</u>

Consider the functions $f_o, f_1, \ldots, f_m$: $E^n \to E^1$ and let $g_i = \max(0, f_i)$ for all $i \in \{1,2,\ldots,m\}$. Consider the problem P: minimize$_x \{f_o(x)$: $f_i(x) \leq 0$, $i=1,2,\ldots,m\}$. Let $\bar{x} \in E^n$ be a feasible solution of problem P and let $I = \{i\colon f_i(\bar{x}) = 0\}$. Furthermore, assume that f_o is locally supportable from below at $\bar{x}$. If $\theta^o(\bar{x}) + \sum_{i \in I} \bar{u}_i \theta^i(\bar{x}) = 0$ where $\theta^o(\bar{x})$ is a subgradient of f_o at $\bar{x}$, $\bar{u}_i \geq 0$ and $\theta^i(\bar{x})$ is a subgradient of g_i at $\bar{x}$ for all $i \in I$, then $\bar{x}$ is a local optimal solution to problem P.

Proof. Since $g_i (i\epsilon I)$ are supportable from below[†] and f_o is supportable from below at $\bar{x}$, then by Theorem 1 we conclude that,

$$f_o(x) \geq f_o(\bar{x}) + \langle \theta^o(\bar{x}), x-\bar{x} \rangle \qquad \text{for all } x\epsilon N_o \tag{26}$$

$$g_i(x) \geq g_i(\bar{x}) + \langle \theta^i(\bar{x}), x-\bar{x} \rangle \qquad \text{for all } x\epsilon N_i, \text{ all } i\epsilon I \tag{27}$$

where $N_i (i\epsilon I\cup\{0\})$ are neighborhoods about $\bar{x}$. Let $N = \min_i\{N_i: i\epsilon I\cup\{0\}\}$. Multiplying (27) by $\bar{u}_i$ and adding (26) and the inequalities corresponding to (27), then

$$\begin{aligned} f_o(x) + \sum_{i\epsilon I} \bar{u}_i g_i(x) &\geq f_o(\bar{x}) + \sum_{i\epsilon I} \bar{u}_i g_i(\bar{x}) \\ &+ \langle \theta^o(\bar{x}) + \sum_{i\epsilon I} \bar{u}_i \theta^i(\bar{x}), x-\bar{x} \rangle \\ &= f_o(\bar{x}) \qquad \text{for all } x\epsilon N \end{aligned} \tag{28}$$

Moreover, if x is a feasible solution, then $g_i(x) = 0$ for $i=1,2,\ldots,m$. This in addition to (28) implies the result.

Corollary

Under the same hypotheses of the above theorem, if f_o is pseudo-convex and f_i $(i=1,2,\ldots,m)$ are quasi-convex, then $\bar{x}$ solves problem P.

Proof. By quasi-convexity of f_i for all $i\epsilon\{1,2,\ldots,m\}$ it follows

---

[†]The construction of $g_i(i\epsilon I)$ imply their supportability from below. Notice that no supportability of $f_i(i\epsilon I)$ is required.

that $\{x: f_i(x) \le 0, i=1,2,\dots,m\}$ is a convex set. Since f_o is pseudoconvex and the constraint set is convex, then every local minimum is a global minimum. See Mangasarian [34]. This completes the proof.

It may be noticed that the result of the above corollary is an extension of a similar result by Mangasarian [34] when the functions are differentiable.

Since the sufficient condition given by Theorem 16 implicitly implies that $u_o > 0$, we give the following simple sufficient condition which is valid independent of the value of u_o. This condition seems to be more powerful since it only involves subgradients of the objective function. Again the supportability from below at the point under investigation is assumed. Some examples which cannot be investigated by the existing optimality criteria may satisfy the hypotheses of Theorem 17 below. See Appendix B for some examples including the example of the outward cusp due to Kuhn and Tucker.

Theorem 17

Consider the problem P: $\text{minimize}_x\{f_o(x): f_i(x) \le 0, i=1,2,\dots,m\}$, where $f_o,f_1,\dots,f_m: E^n \to E^1$. Let $\bar{x}$ be a feasible solution of the problem and assume that $\theta^o(\bar{x})$ is a subgradient of f_o at $\bar{x}$. If $-\theta^o(\bar{x})$ is a *l.o.n.* to the feasible set at $\bar{x}$, then $\bar{x}$ is a local optimal solution to problem P.

Proof. By Theorem 1 it follows that,

$$f_o(x) \ge f_o(\bar{x}) + \langle\theta^o(\bar{x}),x-\bar{x}\rangle \qquad \text{for all } x\in N_1 \tag{29}$$

where N_1 is some neighborhood about $\bar{x}$. But since $-\theta^{o}(\bar{x})$ is a *l.o.n.* to the feasible set S, then,

$$<-\theta^{o}(\bar{x}),x-\bar{x}> \leq 0 \qquad \text{for all } x \in S \cap N_2 \tag{30}$$

where N_2 is some neighborhood about $\bar{x}$. Let $N = \min(N_1,N_2)$ then (29) and (30) imply that $f_o(x) \geq f_o(\bar{x})$ for all $x \in N \cap S$. This completes the proof.

In this chapter we have discussed various optimality conditions when the assumptions regarding the functions involved are relaxed. The next chapter deals with various duality problems in nonlinear programming.

CHAPTER III

DUALITY IN NONLINEAR PROGRAMMING

By a dual problem one usually means a problem which, in some sense, is closely related to the original (primal) problem and where the optimal solutions of the two problems are equivalent. Duality in linear programming has been well studied and the results have been used in various contexts. During the past few years attention has been focused on duality in nonlinear programming, and a number of different duality formulations and theorems have been developed. However, many of these formulations have apparently no relationship with each other.

The purpose of this chapter is to unify the existing duality formulations, to investigate the relationship between them, and to extend some of the existing duality theorems, particularly when differentiability and/or convexity assumptions are relaxed. We show that existing duality formulations can be derived from the Minmax formulation which is discussed in detail in the following section. The subsequent sections discuss the duality formulations via conjugate functions, via the lagrangian multiplier vector, as well as some other duality formulations. Using the theorems presented in Section 1 we also extend some of the existing duality theorems. Certain geometric and/or economic interpretations of duality are also discussed.

Before proceeding further we adopt the following definition concerning dual programs.

Definition 1

The following two problems

$$P: \text{minimize}_x\{\alpha(x): x \in E\} \text{ where } \alpha: E \to E^1 \text{ and } E \subset E^n.$$

$$D: \text{maximize}_y\{\beta(y): y \in F\} \text{ where } \beta: F \to E^1 \text{ and } F \subset E^m.$$

are said to be *subdual programs* iff they satisfy the bounding property $\alpha(x) \geq \beta(y)$ for all $x \in E$ and $y \in F$. Furthermore, the two problems are said to be *dual programs* iff $\inf_x\{\alpha(x): x \in E\} = \sup_y\{\beta(y): y \in F\}$. In this case the problem P is referred to as the *primal problem* and the problem D as the *dual problem*.

It should be noted that dual programs may have other desirable properties such as,

(a) Close geometrical relationship between them.

(b) Existence and boundedness relationship between solutions to problems P and D.

(c) Symmetric property, i.e. the dual of the dual problem is the primal problem.

1. Minmax Theory and Duality

In this section we approach duality via the well known Minmax theory. Theorem 1 gives conditions under which problems P and D are dual programs. Theorems 2 through 4 deal essentially with boundedness and existence of solutions to the primal and dual problems. These results, which are well known, are used in extending other duality theorems in later sections.

We now present some definitions which are used in this section.

Definition 2

A set $A \subset E^n$ is said to be *compact* iff every open covering of A has a finite subcovering of A.

It can be shown that $A \subset E^n$ is *compact* iff A is closed and bounded. See, for example, [5].

Definition 3

A set K in E^n is said to be a *linear manifold* iff $\lambda x^1 + \mu x^2 \in K$ wherever x^1 and $x^2 \in K$, where λ and μ are any scalars.

Definition 4

A set L in E^n is said to be an *affine manifold* iff $L = K + \{\bar{x}\}$, where K is some linear manifold in E^n and $\bar{x} \in E^n$.

Definition 5

The *relative interior* of a set $A \subset E^n$, denoted by $r(A)$ is the interior of A relative to the minimal affine manifold L that contains A, i.e. $\bar{x} \in r(A)$ iff there exists an $\epsilon > 0$ such that all $x \in L$ satisfying $\|x - \bar{x}\| < \epsilon$ belong to A.

Definition 6

A function f: $E^n \to E^1$ is said to be *lower semi-continuous at* $\bar{x} \in E^n$ iff given $\epsilon > 0$, there corresponds a neighborhood N of $\bar{x}$ such that $f(x) > f(\bar{x}) - \epsilon$ for every $x \in N$. f is said to be *upper semi-continuous at* $\bar{x} \in E^n$ iff $-f$ is lower semi-continuous at $\bar{x}$.

f is said to be *lower semi-continuous (upper semi-continuous)* iff f is lower semi-continuous (upper semi-continuous) at x, for every $x \in E^n$.

The terms lower semi-continuous and upper semi-continuous are abbreviated by *l.s.c.* and *u.s.c.*, respectively.

<u>Definition 7</u>

Consider ϕ: $E \times F \to E^1$ where $E \subset E^n$ and $F \subset E^m$. ϕ is said to be *l.s.c.-u.s.c. on ExF* iff $\phi(.,y)$ is *l.s.c.* on E for any given $y \in F$ and $\phi(x,.)$ is *u.s.c.* on F for any given $x \in E$.

<u>Definition 8</u>

Consider ϕ: $E \times F \to E^1$ where $E \subset E^n$ and $F \subset E^m$. ϕ is said to be *quasi-convex—quasi-concave on ExF* iff $\phi(.,y)$ is quasi-convex on E for any given $y \in F$ and $\phi(x,.)$ is quasi-concave on F for any given $x \in E$.

<u>Definition 9</u>

Consider ϕ: $E \times F \to E^1$ where $E \subset E^n$ and $F \subset E^m$. $(\bar{x},\bar{y}) \in E \times F$ is said to be a *saddle point of ϕ on ExF* iff $\phi(\bar{x},y) \le \phi(\bar{x},\bar{y}) \le \phi(x,\bar{y})$ for all $x \in E$ and $y \in F$.

<u>Definition 10</u>

Consider ϕ: $E \times F \to E^1$ where $E \subset E^n$ and $F \subset E^m$ and consider the following two functions.

$$\alpha(x) = \sup_y \{\phi(x,y) : y \in F\} \qquad \text{for all } x \in E, \text{ and} \tag{1}$$

$$\beta(y) = \inf_x \{\phi(x,y) : x \in E\} \qquad \text{for all } y \in F \tag{2}$$

$y \epsilon F$ is said to be *associated with* $x \epsilon E$ iff $\alpha(x) = \phi(x,y)$, i.e. the $\sup_\xi\{\phi(x,\xi)\colon \xi \epsilon F\}$ is attained at $y \epsilon F$. Similarly $x \epsilon E$ is said to be *associated with* $y \epsilon F$ iff $\beta(y) = \phi(x,y)$, i.e. the $\inf_\xi\{\phi(\xi,y)\colon \xi \epsilon E\}$ is attained at $x \epsilon E$.

From the above two definitions it immediately follows that $(\bar{x},\bar{y})$ is a saddle point of ϕ on ExF iff $\bar{x}$ is associated with $\bar{y}$ and $\bar{y}$ is associated with $\bar{x}$.

Now consider the following two problems.

$$P\colon \text{minimize}_x\{\alpha(x)\colon x \epsilon E\} \tag{3}$$

$$D\colon \text{maximize}_y\{\beta(y)\colon y \epsilon F\} \tag{4}$$

where α and β are as defined above by Equations (1) and (2). It follows immediately that $\alpha(x) \geq \beta(y)$ for all $x \epsilon E$ and $y \epsilon F$. This implies that $\inf_{x \epsilon E} \alpha(x) \geq \sup_{y \epsilon F} \beta(y)$ which may be written in the form $\inf_{x \epsilon E} \sup_{y \epsilon F} \phi(x,y) \geq \sup_{y \epsilon F} \inf_{x \epsilon E} \phi(x,y)$.† Therefore problems P and D are subdual programs. Theorem 1 below gives conditions under which the two problems become dual programs. It also gives some conditions under which the inf and sup are attained. The results of the theorem are well known. See [5], [45] and [47].

---

†It may be noted that $\min_{x \epsilon E} \max_{y \epsilon F} \phi(x,y) \geq \max_{y \epsilon F} \min_{x \epsilon E} \phi(x,y)$ if E and F are compact and if ϕ is *l.s.c.-u.s.c.* on ExF.

Theorem 1

Let $\phi:E\times F\to E^1$ where $E\subset E^n$ and $F\subset E^m$ are both convex sets. Let ϕ be both *l.s.c.-u.s.c.* and quasi-convex—quasi-concave on $E\times F$. Then, $\inf_{x\in E}\sup_{y\in F}\phi(x,y) = \sup_{y\in F}\inf_{x\in E}\phi(x,y)$. Furthermore, if E and F are compact then $\min_{x\in E}\max_{y\in F}\phi(x,y) = \max_{y\in F}\min_{x\in E}\phi(x,y)$.†

In the above theorem the compactness of the sets E and F is needed in order to assert that the inf and the sup are attained. However, this is rather a strong assumption since it implies boundedness of the sets E and F. Hence we present below several theorems on existence and finiteness of the solutions to problems P and D as defined earlier. Theorem 2 that follows is due to Rockafeller [42].

Theorem 2

Let ϕ: $E\times F\to E^1$ where $E\subset E^n$ and $F\subset E^m$ are convex sets. Let ϕ be both convex-concave and *l.s.c.-u.s.c.* on $E\times F$. Let $\alpha(x) = \sup_y\{\phi(x,y): y\in F\}$ for all $x\in E$ and $\beta(y) = \inf_x\{\phi(x,y): x\in E\}$ for all $y\in F$. If condition (i) below is satisfied, then $\max_{y\in F}\beta(y) = \inf_{x\in E}\alpha(x) < \infty$, and if condition (ii) below is satisfied, then $\sup_{y\in F}\beta(y) = \min_{x\in E}\alpha(x) > -\infty$. Moreover, if both conditions are satisfied, then ϕ has a saddle point on $E\times F$.

Condition (i). No nonzero vector y^o has the property that for all $y\in r(F)$ and $x\in r(E)$ the ray $\{y + \lambda y^o: \lambda\geq 0\}$ is contained in F, and $\phi(x,y + \lambda y^o)$ is a nonzero decreasing function of $\lambda\geq 0$.

---

†It can be shown that if E and F are compact and ϕ is *l.s.c.-u.s.c.* on $E\times F$, then $\min_{x\in E}\max_{y\in F}\phi(x,y) = \max_{y\in F}\min_{x\in E}\phi(x,y)$ iff ϕ has a saddle point on $E\times F$. See [45] for example.

<u>Condition (ii)</u>. No nonzero vector x^o has the property that for all $y \epsilon r(F)$ and $x \epsilon r(E)$ the ray $\{x + \lambda x^o: \lambda \geq 0\}$ is contained in E, and $\phi(x+\lambda x^o, y)$ is a nonincreasing function of $\lambda \geq 0$.

We consider below two alternative theorems on the existence of solutions to problems P and D. The first theorem uses the notion of the high and low value properties of ϕ which Mangasarian and Ponstein [36] defined as a simplification of the B-property introduced by Stoer [48].

<u>Definition 11</u>

Let ϕ: $ExF \to E^1$ where $E \subset E^n$ and $F \subset E^m$, and assume that ϕ is *l.s.c.-u.s.c.* on ExF. ϕ is said to have the *high value property at* $(\bar{x},\bar{y}) \epsilon ExF$ iff there exists a closed neighborhood N about $\bar{y}$ and a compact convex set $A \subset E$ such that $\phi(\bar{x},\bar{y}) \geq \max_{y \epsilon F \cap N} \min_{x \epsilon A} \phi(x,y)$. Similarly, ϕ is said to have the *low value property at* $(\bar{x},\bar{y}) \epsilon ExF$ iff there exists a closed neighborhood N of $\bar{x}$ and a compact convex set $B \subset F$ such that $\phi(\bar{x},\bar{y}) \leq \min_{x \epsilon E \cap N} \max_{y \epsilon B} \phi(x,y)$.

The following theorem is due to Mangasarian and Ponstein [36]. It gives necessary and sufficient conditions that a solution of one problem exists given that a solution of the other problem exists.

<u>Theorem 3</u>

Let ϕ: $ExF \to E^1$ be continuous and convex-concave on ExF, where E and F are closed and convex. Let $x^1 \epsilon E$ solve problem P as defined by (3) and assume that $y^1 \epsilon F$ is associated with x^1. Then there exists $y^o \epsilon F$ such that y^o solves problem D as defined by (4) and (x^1,y^o) is a saddle point† of ϕ on ExF iff ϕ has the low value property at (x^1,y^1). Let $y^2 \epsilon F$ solve problem D and assume that $x^2 \epsilon E$ is associated with y^2. Then there exists $x^o \epsilon E$ such that x^o solves problem P and (x^o,y^2) is a saddle point of ϕ on ExF iff ϕ has the high value property at (x^2,y^2).

The following theorem due to Roode [45] gives a slight generalization of a similar theorem by Mangasarian and Ponstein [36]. The theorem represents a substitute for the above theorem (in the sufficiency sense) where the low and high value properties are substituted with strict quasi-concavity of $\phi(x^1,.)$ in a neighborhood of y^1, and strict quasi-convexity of $\phi(.,y^2)$ in a neighborhood of x^2, respectively, where x^1 solves problem P (y^1 associated with x^1) and y^2 solves problem D (x^2 associated with y^2).

Theorem 4

Let ϕ: $ExF \to E^1$ be *l.s.c.-u.s.c.* on ExF, where $E \subset E^n$ and $F \subset E^m$ are both closed and convex. Furthermore assume that ϕ is *s*-strictly quasi-convex—*s*-strictly quasi-concave†† on ExF. Let $x^1 \epsilon E$ solve problem P as defined by Equation (3) and assume that $y^1 \epsilon F$ is associated with x^1. If $\phi(x^1,.)$ is strictly quasi-concave in some neighborhood of y^1, then y^1 solves problem D as defined by Equation (4), and (x^1,y^1) is a saddle point of ϕ on ExF. Now let $y^2 \epsilon F$ solve problem D and assume that $x^2 \epsilon E$

---

†Recall that if (x^1,y^o) is a saddle point of ϕ on ExF, then x^1 is associated with y^o and y^o is associated with x^1. It then follows that $\phi(x^1,y^o) = \phi(x^1,y^1)$.

††ϕ: $ExF \to E^1$ is *s*-strictly quasi-convex—*s*-strictly quasi-concave on ExF iff $\phi(.,y)$ is *s*-strictly quasi-convex on E for every $y \epsilon F$ and $\phi(x,.)$ is *s*-strictly quasi-concave on F for every $x \epsilon E$. See Definition 14, Chapter II.

is associated with y^2. If $\phi(.,y^2)$ is strictly quasi-convex in some neighborhood of x^2, then x^2 solves problem P and (x^2,y^2) is a saddle point of ϕ on ExF.

2. Duality via Conjugate Functions

In this section we first introduce conjugate functions due to Fenchel [18] and then present his important duality theorem. We then show that Fenchel's dual formulation can be obtained as a special case from the Minmax formulation discussed earlier. Application of Theorems 2 through 4 then permits us to partially relax convexity assumptions and to consider slightly different versions of Fenchel's theorem. We also give alternative ways by which a nonlinear programming problem can be expressed in the primal form of Fenchel's formulation.

Sup and Inf Conjugate Functions

Fenchel introduced the notion of conjugate functions [18]. Not until recently has this useful concept been fruitfully utilized in the field of nonlinear programming. The concept of conjugate functions is closely related to the notion of outernormals and subgradients. This relationship is highlighted in this section. It should be noted that Fenchel adopted conjugate functions when the original functions are either convex or concave. The definitions given below are not restricted to convex and concave functions. Our definition of the sup conjugate (inf conjugate) is identical to Fenchel's definition when the original function is convex (concave).

Definition 11

Let $g: E \to E^1$ where $E \subset E^n$. $g^*: E^* \to E^1$ is said to be the *sup conjugate* of g iff $g^*(x^*) = \sup_x\{<x^*,x> - g(x): x \in E\}$ for all $x^* \in E^*$, where $E^* = \{x^*: x^* \in E^n, \sup_x\{<x^*,x> - g(x): x \in E\} < \infty\}$.

It is interesting to note that E^* is a convex set and g^* is a convex function regardless of E and g. This result is given by the following remark.

Remark 1

Let $g: E \to E^1$ where $E \subset E^n$. Then,

$$E^* = \{x^*: x^* \in E^n, \sup_x\{<x^*,x> - g(x): x \in E\} < \infty\}$$

is a convex set, and $g^*: E^* \to E^1$ where $g^*(x^*) = \sup_x\{<x^*,x> - g(x): x \in E\}$ is a convex function.

Proof. Let x^* and $\bar{x}^* \in E^*$. Then we want to show that $\lambda x^* + (1-\lambda)\bar{x}^* \in E^*$ for all $\lambda \in (0,1)$. Consider $\lambda x^* + (1-\lambda)\bar{x}^*$ for any $\lambda \in (0,1)$, then,

$$\sup_x\{<\lambda x^* + (1-\lambda)\bar{x}^*,x> - g(x): x \in E\} \leq \lambda \sup_x\{<x^*,x> - g(x): x \in E\}$$

$$+ (1-\lambda) \sup_x\{<\bar{x}^*,x> - g(x): x \in E\} < \infty.$$

Therefore $\lambda x^* + (1-\lambda)\bar{x}^* \in E^*$ and hence E^* is convex. We now show that g^* is a convex function. It follows immediately from the above inequality that $g^*(\lambda x^* + (1-\lambda)\bar{x}^*) \leq \lambda g^*(x^*) + (1-\lambda)g^*(\bar{x}^*)$ for any x^*

and $\bar{x}^* \epsilon E^*$ and $\lambda\epsilon(0,1)$. This shows convexity of g^* and the proof is complete.

We may consider the following interpretation of $g^*(x^*)$. By definition, for a given $x^* \epsilon E^*$ it is required, in some sense, to maximize $\langle x^*,x\rangle - g(x)$ over $x\epsilon E$. In other words we want to maximize the distance along the y-axis between the hyperplane given by $y = \langle x^*,x\rangle$ and the surface $y = g(x)$.† If the maximum occurs at $\bar{x}\epsilon E$, as in Figure 1, then $g^*(x^*) = \langle x^*,\bar{x}\rangle - g(\bar{x})$. Hence if the hyperplane is moved parallel to itself until it supports the epigraph of g at $(\bar{x},g(\bar{x}))$, then its intercept on the y-axis is $-g^*(x^*)$. This is obvious from the figure. It may also be noted that $(x^*,-1)$ is an *o.n.* to the epigraph of g at $(\bar{x},g(\bar{x}))$.

The above discussion is made precise by the following remark. We assert that the sup in Definition 11 is attained at $\bar{x}$ when $x^* \epsilon E^*$ is given iff $(x^*,-1)$ is an *o.n.* to the epigraph of g at $(\bar{x},g(\bar{x}))$. It may be recalled from Theorem 1 in Chapter II that this is equivalent to the supportability from below of g at $\bar{x}$.

Remark 2

$g^*(x^*) = \langle x^*,\bar{x}\rangle - g(\bar{x})$ for some $\bar{x}\epsilon E$ iff $(x^*,-1)\epsilon E^{n+1}$ is an *o.n.* to the epigraph of g at $(\bar{x},g(\bar{x}))$.

Proof. First assume that $(x^*,-1)$ is an *o.n.* to the epigraph of

---

†The direction of the distance is considered, i.e. if for given $x^* \epsilon E^*$ and $x\epsilon E$, $\langle x^*,x\rangle - g(x) < 0$, the distance between the hyperplane and the surface is negative.

g at $(\bar{x},g(\bar{x}))$. It immediately follows that $\langle x^*,x\rangle - g(x) \le \langle x^*,\bar{x}\rangle - g(\bar{x})$ for all $x\epsilon E$. This further implies that $\langle x^*,\bar{x}\rangle - g(\bar{x}) = \sup_x\{\langle x^*,x\rangle - g(x)\colon x\epsilon E\} = g^*(x^*)$. On the other hand if $g^*(x^*) = \langle x^*,\bar{x}\rangle - g(\bar{x})$ for some $\bar{x}\epsilon E$, then $g(x) \ge g(\bar{x}) + \langle x^*,x-\bar{x}\rangle$ for all $x\epsilon E$ and hence $(x^*,-1)$ is an *o.n.* to the epigraph of g at $(\bar{x},g(\bar{x}))$. This completes the proof.

It may be noted that in the above remark x^* is a subgradient of g at $\bar{x}$. We now consider the following definition of the inf conjugate function.

Definition 12

Let $h\colon F\to E^1$ where $F\subset E^n$. $h_*\colon F^*\to E^1$ is said to be the *inf conjugate* of h iff $h_*(x^*) = \inf_x\{\langle x^*,x\rangle - h(x)\colon x\epsilon F\}$ for all $x^*\epsilon F^*$ where $F^* = \{x^*\colon x^*\epsilon E^n, \inf_x\{\langle x^*,x\rangle - h(x)\colon x\epsilon F\} > -\infty\}$.

It is interesting to note that F^* is a convex set and h_* is a concave function regardless of F and h. The following remark states this result. The proof is similar to that of Remark 1 and is omitted.

Remark 3

Let $h\colon F\to E^1$ where $F\subset E^n$. Then

$$F^* = \{x^*\colon x^*\epsilon\ E^n, \inf_x\{\langle x^*,x\rangle - h(x)\colon x\epsilon F\} > -\infty\}$$

is a convex set, and $h_*\colon F^*\to E^1$ where $h_*(x^*) = \inf_x\{\langle x^*,x\rangle - h(x)\colon x\epsilon F\}$ is a concave function.

An interpretation of $h_*(x^*)$ similar to that of $g^*(x^*)$ can be given as follows. For a given x^* it is required to minimize

$\langle x^*,x\rangle - h(x)$ for all $x\epsilon F$. We therefore maximize the distance between the surface $y = h(x)$ and the hyperplane $y = \langle x^*,x\rangle$ along the y-axis. If the maximum is attained at $\bar{x}\epsilon F$ as in Figure 2, then $h_*(x^*) = \langle x^*,\bar{x}\rangle -h(\bar{x})$. Hence if the hyperplane is moved vertically and parallel to itself such that it supports the hypograph of h at $(\bar{x},h(\bar{x}))$, then the intercept of the hyperplane on the y-axis gives $-h_*(x^*)$.

The following remark relates $(-x^*,1)$ to the hypograph of h at $(\bar{x},h(\bar{x}))$. The proof of the remark is similar to that of Remark 2 and is omitted.

<u>Remark 4</u>

$h_*(x^*) = \langle x^*,\bar{x}\rangle - h(\bar{x})$ for some $\bar{x}\epsilon F$ iff $(-x^*,1)$ is an *o.n.* to the hypograph of h at $(\bar{x},h(\bar{x}))$.

It is interesting to note that the conjugate of the conjugate of certain convex (concave) functions are the original functions. This fact will be used later and hence is stated as Remark 5 below. For a proof one may refer to Rockafeller [44] or Luenberger [32].

<u>Remark 5</u>

Let $g: E\to E^1$ be a closed† convex function where $E\subset E^n$ is a convex set. Then $g^{**} = g$ and $E^{**} = E$. Similarly let $h: F\to E^1$ be a closed concave function where F is a convex set, then $h_{**} = h$ and $F^{**} = F$.

<u>Fenchel's Duality Theorem and Extensions</u>

Fenchel [18] developed an important duality theorem which is

---

†A convex function g defined on a convex set E is said to be *closed* iff for any $\bar{x}\epsilon E$ where $\underline{\lim}_{x\to\bar{x}}\, g(x)$ exists, $g(\bar{x}) = \underline{\lim}_{x\to\bar{x}}\, g(x)$. It can be shown that g is closed iff it is lower semi-continuous. See Rockafeller [44].

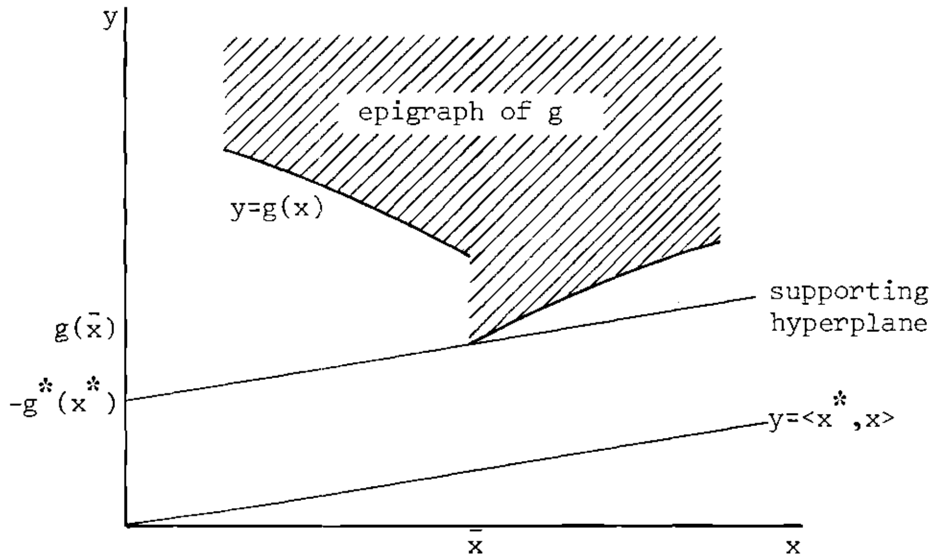

Figure 1. An Interpretation of the Sup Conjugate Function

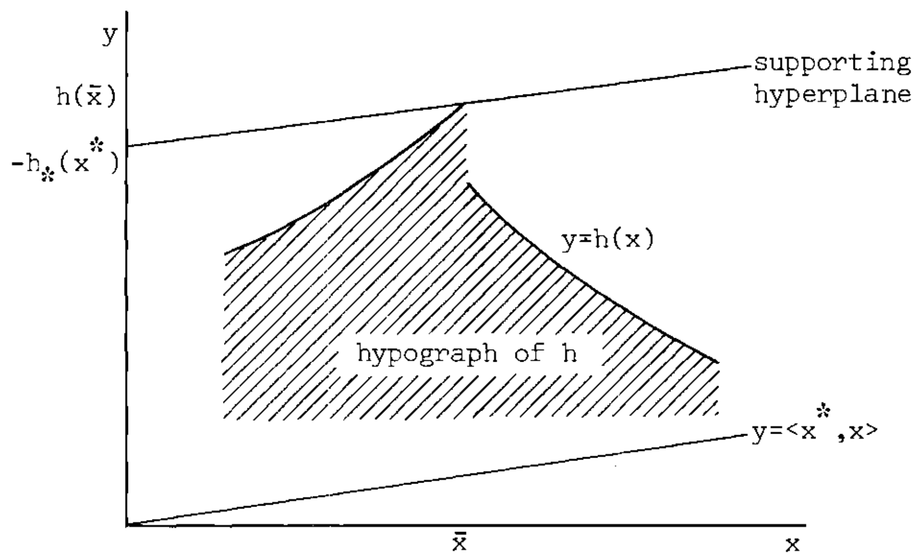

Figure 2. An Interpretation of the Inf Conjugate Function

based on the conjugate functions discussed above. He considered the following problems:

$$P:\ \text{minimize}_x\{g(x)-h(x):\ x\epsilon E\cap F\},\quad \text{and}$$

$$D:\ \text{maximize}_{x^*}\{h_*(x^*)-g^*(x^*):\ x^*\epsilon E^*\cap F^*\}\ \text{where,}$$

$g:\ E\to E^1$, $h:\ F\to E^1$, E and F are convex sets in E^n. Moreover, g is assumed to be closed and convex whereas h is closed and concave. g^*, h_*, E^*, and F^* are given by Definitions 11 and 12.

The following remark shows that the problems P and D above are subdual programs. It should be noticed that this is true without any restrictions on the functions or the sets involved. Imposing the restrictions given above, the two problems form two dual programs. This result due to Fenchel is given below as Theorem 5. For a proof one may refer to [19], [29], and [32].

Remark 6

$g(x) - h(x) \geq h_*(x^*) - g^*(x^*)$ for every $x\epsilon E\cap F$ and $x^*\epsilon E^*\cap F^*$, where g, h, g^*, h_*, E, F, E^*, and F^* are as in Definitions 11 and 12.

Proof. By definition the following two inequalities hold.

$$h_*(x^*) = \inf_x\{\langle x^*,x\rangle - h(x):\ x\epsilon F\} \leq \langle x^*,x\rangle - h(x) \qquad \text{all } x\epsilon F,\ \text{all } x^*\epsilon F^*.$$

$$-g^*(x^*) = -\sup_x\{\langle x^*,x\rangle - g(x):\ x\epsilon E\} = \inf_x\{g(x) - \langle x^*,x\rangle:\ x\epsilon E\}$$

$$\leq g(x) - \langle x^*,x\rangle \qquad \text{all } x\epsilon E,\ \text{all } x^*\epsilon E^*$$

Adding the above inequalities for any $x\epsilon E\cap F$ and $x^*\epsilon E^*\cap F^*$ the result follows immediately.

Theorem 5

Let $g: E\to E^1$ be a closed convex function and $h: F\to E^1$ be a closed concave function where E and F are both convex sets in E^n. Consider the following two cases.

(i) $\inf_x\{g(x) - h(x): x\epsilon E\cap F\} > -\infty$, and $r(E)\cap r(F) \neq \phi$.

(ii) $\sup_{x^*}\{h_*(x^*) - g^*(x^*): x^*\epsilon E^*\cap F^*\} < \infty$, and $r(E^*)\cap r(F^*) \neq \phi$.

where g^*, E^*, h_*, and F^* are as in Definitions 11 and 12. If condition (i) is satisfied, then there exists an $\bar{x}^*\epsilon E^*\cap F^*$ such that,

$$\inf_x\{g(x)-h(x): x\epsilon E\cap F\} = \max_{x^*}\{h_*(x^*)-g^*(x^*): x^*\epsilon E^*\cap F^*\} = h_*(\bar{x}^*) - g^*(\bar{x}^*).$$

If the inf on the left is achieved by some $\bar{x}\epsilon E\cap F$, then,

$$\max_x\{<\bar{x}^*,x> - g(x): x\epsilon E\} = <\bar{x}^*,\bar{x}> - g(\bar{x}), \text{ and}$$

$$\min_x\{<\bar{x}^*,x> - h(x): x\epsilon F\} = <\bar{x}^*,\bar{x}> - h(\bar{x}).$$

If condition (ii) is satisfied, then there exists an $x^o\epsilon E\cap F$ such that,

$$\sup_{x^*}\{h_*(x^*)-g^*(x^*): x^*\epsilon E^*\cap F^*\} = \min_x\{g(x)-h(x): x\epsilon E\cap F\} = g(x^o) - h(x^o).$$

We may replace the assumptions $\inf_x\{g(x)-h(x): x\epsilon E\cap F\} > -\infty$ and $\sup_{x^*}\{h_*(x^*)-g^*(x^*): x^*\epsilon E^*\cap F^*\} < \infty$ in the above theorem by $E^*\cap F^* \neq \phi$ and $E\cap F \neq \phi$, respectively.

We now show that Fenchel's formulation can be obtained as a special case from the Minmax formulation. This is done via two different choices of the function ϕ. In the first case we require that h is concave and in the second case we require that g is convex. Besides giving more flexibility, this leads to more general results.

We first consider the function ϕ: $E \times F^{*\dagger} \to E^1$ where E and F are convex sets in E^n. Let $\phi(x,x^*) = g(x) + h_*(x^*) - \langle x^*,x\rangle$ for all $x \in E$ and $x^* \in F^*$, where g: $E \to E^1$ and h_*: $F^* \to E^1$. We further assume that h is closed and concave. We then construct the functions α and β as follows.

$$\alpha(x) = \sup_{x^*}\{g(x) + h_*(x^*) - \langle x^*,x\rangle : x^* \in F^*\}$$

$$= g(x) - \inf_{x^*}\{\langle x^*,x\rangle - h_*(x^*) : x^* \in F^*\}$$

$$= g(x) - h_{**}(x) \qquad \text{if } x \in E \cap F^{**}$$

$$\infty \qquad \text{if } x \in E \cap (F^{**})^c$$

where $F^{**} = \{x: \inf_{x^*}\{\langle x^*,x\rangle - h_*(x^*): x^* \in F^*\} > -\infty\}$, and

$$h_{**}(x) = \inf_{x^*}\{\langle x^*,x\rangle - h_*(x^*): x^* \in F^*\} \qquad \text{for } x \in F^{**}$$

It follows from Remark 5 that $F^{**} = F$ and $h_{**} = h$. Similarly we

---

†Note that $F^* = \{x^*: \inf_x\{\langle x^*,x\rangle - h(x): x \in F\} > -\infty\}$ is a convex set.

construct the function β as follows.

$$\beta(x^*) = \inf_x\{g(x) + h_*(x^*) - \langle x^*,x\rangle : x\epsilon E\}$$

$$= h_*(x^*) - \sup_x\{\langle x^*,x\rangle - g(x): x\epsilon E\}$$

$$= h_*(x^*) - g^*(x^*) \qquad \text{if } x^*\epsilon E^*\cap F^*$$

$$-\infty \qquad \text{if } x^*\epsilon (E^*)^c\cap F^*$$

where $E^* = \{x^*: \sup_x\{\langle x^*,x\rangle - g(x): x\epsilon E\} < \infty\}$, and

$$g^*(x^*) = \sup_x\{\langle x^*,x\rangle - g(x): x\epsilon E\} \qquad \text{for } x^*\epsilon E^*$$

Therefore the problems P and D to minimize$_{x\epsilon E\cap F}$ $\alpha(x)$ and to maximize$_{x^*\epsilon E^*\cap F^*}$ $\beta(x^*)$ become,

$$\text{P: minimize}_x\{g(x)-h(x): x\epsilon E\cap F\} \tag{5}$$

$$\text{D: maximize}_{x^*}\{h_*(x^*) - g^*(x^*): x^*\epsilon E^*\cap F^*\}. \tag{6}$$

This is precisely the form of duality given by Fenchel. It should be emphasized that no convexity of g is assumed.

On the other hand, by choosing ϕ: $F\times E^{*\dagger} \to E^1$ where F and E are

---

†Note that $E^* = \{x^*: \sup_x\{\langle x^*,x\rangle - g(x): x\epsilon E\} < \infty\}$ is a convex set.

convex sets in E^n and $\phi(x,x^*) = -h(x) - g^*(x^*) + \langle x^*,x\rangle$ for $x\epsilon F$ and $x^*\epsilon E^*$, we need only assume that g is closed and convex function to obtain Fenchel's formulation. We calculate α and β as follows.

$$\alpha(x) = \sup_{x^*}\{-h(x) - g^*(x^*) + \langle x^*,x\rangle : x^*\epsilon E^*\}$$

$$= -h(x) + g^{**}(x) \qquad \text{if } x\epsilon E^{**}\cap F$$

$$\infty \qquad \text{if } x\epsilon (E^{**})^c\cap F$$

where $$E^{**} = \{x: \sup_{x^*}\{\langle x^*,x\rangle - g^*(x^*)\} < \infty\}, \text{ and}$$

$$g^{**}(x) = \sup_{x^*}\{\langle x^*,x\rangle - g^*(x^*): x^*\epsilon E^*\} \quad \text{for } x\epsilon E^{**}$$

It follows from Remark 5 that $g^{**} = g$ and $E^{**} = E$.

$$\beta(x^*) = \inf_x\{-h(x) - g^*(x^*) + \langle x^*,x\rangle : x\epsilon F\}$$

$$= -g^*(x^*) + h_*(x^*) \qquad \text{if } x^*\epsilon E^*\cap F^*$$

$$-\infty \qquad \text{if } x^*\epsilon E^*\cap (F^*)^c$$

where $$F^* = \{x^*: \inf_x\{\langle x^*,x\rangle - h(x): x\epsilon F\} > -\infty\}$$

$$h_*(x^*) = \inf_x\{\langle x^*,x\rangle - h(x): x\epsilon F\} \quad \text{for } x^*\epsilon F^*.$$

Hence Fenchel's formulation follows. Recall that we assumed closed convexity of g^* but no restrictions on h are required.

We have earlier established the bounding property of problems P and D. Some extra assumptions (as in Fenchel's theorem) are needed to show the equivalence of the optimal solutions of the two problems. We will now specialize the function ϕ as above, in Theorems 1 through 4, to obtain the primal and dual problems of Fenchel. This leads to Theorems 6 through 9 which represent different versions of Fenchel's theorem and some extensions of his results. In Theorems 6 through 9 below, E^*, F, g^*, and h are refined as follows:

$$E^* = \{x^*: \sup_x\{<x^*,x> - g(x): x\epsilon E\} < \infty\}. \tag{7}$$

$$F = \{x: \inf_{x^*}\{<x^*,x> - h_*(x^*): x^*\epsilon F^*\} > -\infty\}. \tag{8}$$

$$g^*(x^*) = \sup_x\{<x^*,x> - g(x): x\epsilon E\} \qquad \text{for all } x^*\epsilon E^*. \tag{9}$$

$$h(x) = \inf_{x^*}\{<x^*,x> - h_*(x^*): x^*\epsilon F^*\} \quad \text{for all } x\epsilon F. \tag{10}$$

<u>Theorem 6</u>

Let $\phi(x,x^*) = g(x) + h_*(x^*) - <x^*,x>$ for all $x\epsilon E$ and $x^*\epsilon F^*$ where E and F^* are closed convex sets in E^n. Suppose that g: $E \to E^1$ is *l.s.c.*, h_*: $F^* \to E^1$ is *u.s.c.* and concave, and $\phi(.,x^*)$ is quasi-convex on E for each $x^*\epsilon F^*$. Then by Theorem 1,

$$\inf_x\{g(x) - h(x): x\epsilon E\cap F\} = \sup_{x^*}\{h_*(x^*) - g^*(x^*): x^*\epsilon E^*\cap F^*\}$$

where E^*, F, g^*, and h are defined by Equations (7) through (10). Moreover, if E and F^* are bounded, then the inf and sup above are attained.

<u>Theorem 7</u>

Let $\phi(x,x^*) = g(x) + h_*(x^*) - \langle x^*,x\rangle$ for all $x\epsilon E$ and $x^*\epsilon F^*$ where E and F^* are closed convex sets in E^n. Suppose that $g: E\to E^1$ is *l.s.c.* and convex and $h_*: F^*\to E^1$ is *u.s.c.* and concave. If condition (i) below is satisfied, then by Theorem 2,

$$\max_{x^*}\{h_*(x^*) - g^*(x^*): x^*\epsilon E^*\cap F^*\} = \inf_x\{g(x) - h(x): x\epsilon E\cap F\} < \infty.$$

If condition (ii) below is satisfied, then by Theorem 2,

$$\min_x\{g(x) - h(x): x\epsilon E\cap F\} = \sup_{x^*}\{h_*(x^*) - g^*(x^*): x^*\epsilon E^*\cap F^*\} > -\infty.$$

If both conditions are satisfied then,

$$\min_x\{g(x) - h(x): x\epsilon E\cap F\} = \max_{x^*}\{h_*(x^*) - g^*(x^*): x^*\epsilon E^*\cap F^*\}$$

where E^*, F, g^*, and h are defined by Equations (7) through (10).

<u>Condition (i)</u>. No nonzero vector $\bar{x}^*$ has the property that for all $x^*\epsilon r(F^*)$ and $x\epsilon r(E)$, the ray $\{x^* + \lambda\bar{x}^*: \lambda \geq 0\}$ is contained in F^* and $\phi(x,x^* + \lambda\bar{x}^*)$ is a nonzero decreasing function of $\lambda \geq 0$.

<u>Condition (ii).</u> No nonzero vector $\bar{x}$ has the property that for all $x^*\epsilon r(F^*)$ and $x\epsilon r(E)$, the ray $\{x + \lambda\bar{x}: \lambda \geq 0\}$ is contained in E, and $\phi(x + \lambda\bar{x}, x^*)$ is a nonincreasing function of $\lambda \geq 0$.

<u>Theorem 8</u>

Let $\phi(x,x^*) = g(x) + h_*(x^*) - \langle x^*,x\rangle$ for all $x\epsilon E$ and $x^*\epsilon F^*$ where E and F^* are closed convex sets in E^n. Suppose that $g: E \to E^1$ is continuous and convex and $h_*: F^* \to E^1$ is continuous and concave. Assume that $\bar{x}$ solves problem P given by Equation (5) and $\bar{x}^*$ is associated with $\bar{x}$. Then by Theorem 3, there exists an $\overset{o}{x}{}^*$ that solves problem D given by Equation (6), and $(\bar{x},\overset{o}{x}{}^*)$ is a saddle point of ϕ on ExF^* iff ϕ has the low value property at $(\bar{x},\bar{x}^*)$. Further suppose that $\hat{x}^*$ solves problem D and $\hat{x}$ is associated with $\hat{x}^*$. Then by Theorem 3 there exists an x^o that solves problem P and $(x^o,\hat{x}^*)$ is a saddle point of ϕ on ExF^* iff ϕ has the high value property at $(\hat{x},\hat{x}^*)$.

It may be noticed that if in the above theorem $(\bar{x},\overset{o}{x}{}^*)$ is a saddle point of ϕ on ExF^*, then

$$g(\bar{x}) + h_*(x^*) - \langle x^*,\bar{x}\rangle \leq g(\bar{x}) + h_*(\overset{o}{x}{}^*) - \langle \overset{o}{x}{}^*,\bar{x}\rangle$$

$$\leq g(x) + h_*(\overset{o}{x}{}^*) - \langle \overset{o}{x}{}^*,x\rangle \qquad \text{all } x\epsilon E, \text{ and } x^*\epsilon F^*.$$

By rearranging the terms in the above inequalities it follows that,

$$g(x) \geq g(\bar{x}) + \langle \overset{o}{x}{}^*,x-\bar{x}\rangle \qquad \text{for all } x\epsilon E, \text{ and}$$

$$h_*(x^*) \le h_*(\overset{o}{x}{}^*) + \langle \bar{x}, x^* - \overset{o}{x}{}^* \rangle \qquad \text{for all } x^* \epsilon F^*$$

This shows that $\overset{o}{x}{}^*$ is a subgradient of g at $\bar{x}$ and $\bar{x}$ is a subgradient of h_* at $\overset{o}{x}{}^*$. This is an interesting relationship because it states that the optimal solution of problem D is a subgradient of g at the optimal solution of problem P. Also the optimal solution of problem P is a subgradient of h_* at the optimal solution of problem D. This again highlights the strong existing duality relationship between problems P and D. Furthermore, if $(\bar{x},\overset{o}{x}{}^*)$ is a saddle point, then $\bar{x}$ is associated with $\overset{o}{x}{}^*$ and $\overset{o}{x}{}^*$ is associated with $\bar{x}$. This along with the hypothesis that $\bar{x}^*$ is associated with $\bar{x}$ imply that $h_*(\bar{x}^*) = h_*(\overset{o}{x}{}^*) + \langle \bar{x}, \bar{x}^* - \overset{o}{x}{}^* \rangle$. By concavity of h_* it follows that either $\bar{x}^* = \overset{o}{x}{}^*$, or h_* is linear on the segment line joining $\bar{x}^*$ and $\overset{o}{x}{}^*$. Similar interpretation can be given to the second part of Theorem 8 when $(x^o, \hat{x}^*)$ is a saddle point of ϕ on $E \times F^*$.

<u>Theorem 9</u>

Let $\phi(x,x^*) = g(x) + h_*(x^*) - \langle x^*, x \rangle$ for all $x \epsilon E$ and $x^* \epsilon F^*$ where E and F^* are closed convex sets in E^n. Suppose that $g: E \to E^1$ is *l.s.c.* and $h_*: F^* \to E^1$ is *u.s.c.* and concave, and assume that $\phi(.,x^*)$ is *s*-strictly quasi-convex on E for each $x^* \in F^*$. Let $\bar{x}$ solve problem P given by Equation (5) and assume that $\bar{x}^*$ is associated with $\bar{x}$. Then by Theorem 4, if h_* is strictly concave in some neighborhood of $\bar{x}^*$, then $\bar{x}^*$ solves problem D given by Equation (6), and $(\bar{x},\bar{x}^*)$ is a saddle point of ϕ on $E \times F^*$. Further suppose that $\hat{x}^*$ solves problem D, and $\hat{x}$ is associated with $\hat{x}^*$, then by Theorem 4 if $\phi(.,\hat{x}^*)$ is strictly quasi-convex in some neighborhood of $\hat{x}$, it follows that $\hat{x}$

solves problem P and $(\hat{x},\hat{x}^*)$ is a saddle point of ϕ on $E\times F^*$.

Results similar to those of Theorems 6 through 9 above can be obtained by letting $\phi(x,x^*) = -h(x) - g^*(x^*) + \langle x^*,x\rangle$ for all $x\epsilon F$ and $x^*\epsilon E^*$, and assuming that E and F are convex sets in E^n and g is closed and convex.

It should be noted that in Theorem 5 (Fenchel's theorem) above and also in Theorems 6 through 9 some sort of convexity assumption was required. However, the convexity assumption can be replaced by a weaker assumption, namely the existence of a separating hyperplane between the hypograph of h and the set $A = \{(x,y): x\epsilon E,\ y\geq g(x)-\mu\}$ where $\mu = \inf_x\{g(x)-h(x): x\epsilon E\cap F\}$. In the following theorem we assume that a solution to the problem P given by Equation (5) exists in addition to some other hypotheses which insure the existence of a separating hyperplane.

Theorem 10

Let $g: E\to E^1$ and $h: F\to E^1$ where E and F are subsets of E^n such that $\mathrm{int}(E\cap F) \neq \phi$. Let $\mu = \inf_x\{g(x)-h(x): x\epsilon E\cap F\}$ be finite and attained at some $\bar{x}\epsilon E\cap F$. Furthermore assume that $(\bar{x},h(\bar{x}))\epsilon(\partial[A])\cap(\partial[B])$ and $(\mathrm{int}[A])\cap(\mathrm{int}[B]) = \phi$, where $A = \{(x,y): x\epsilon E,\ y\geq g(x)-\mu\}$ and $B = \{(x,y): x\epsilon F,\ y<h(x)\}$. Then

$$\mu = \min_x\{g(x)-h(x): x\epsilon E\cap F\} = \max_{x^*}\{h_*(x^*)-g^*(x^*): x^*\epsilon E^*\cap F^*\}.$$

Proof. By Remark 6 it follows that $g(x)-h(x) \geq h_*(x^*)-g^*(x^*)$ for all $x\epsilon E\cap F$ and $x^*\epsilon E^*\cap F^*$. Therefore it suffices to show that there exists an $x^*\epsilon E^*\cap F^*$ such that $\mu = h_*(x^*) - g^*(x^*)$. But since

$(\mathrm{int}[A])\cap(\mathrm{int}[B]) = \phi$ and $(\bar{x},h(\bar{x}))\epsilon(\partial[A])\cap(\partial[B])$, then there exists a hyperplane H passing through $(\bar{x},h(\bar{x}))$ which separates [A] and [B]. Let $H = \{(x,y)\colon x\epsilon E^n,\ y\epsilon E^1,\ \langle p,(x,y)-(\bar{x},h(\bar{x}))\rangle=0\}$ where p is a nonzero vector in E^{n+1}. $p_{n+1} \neq 0$ because if not the hyperplane H would separate E and F contradicting the hypothesis $\mathrm{int}(E\cap F) \neq \phi$. Hence without loss of generality let $p = (x^*,-1)$ where $x^*\epsilon E^n$. Therefore if $(x,y)\epsilon H$, then $y = h(\bar{x}) + \langle x^*,x-\bar{x}\rangle$. Since H supports both [A] and [B] at $(\bar{x},h(\bar{x}))$, then it follows that,

$$0 = \inf_x\{g(x)-\mu-y\colon x\epsilon E,\ (x,y)\epsilon H\} = \sup_x\{h(x)-y\colon x\epsilon F,\ (x,y)\epsilon H\}.$$

This implies that,

$$\begin{aligned} 0 &= \inf_x\{g(x)-\mu-h(\bar{x})-\langle x^*,x-\bar{x}\rangle\colon x\epsilon E\} \\ &= -\sup_x\{\langle x^*,x\rangle-g(x)\colon x\epsilon E\} -\mu-h(\bar{x}) \\ &\quad + \langle x^*,\bar{x}\rangle = -g^*(x^*)-\mu-h(\bar{x}) + \langle x^*,\bar{x}\rangle, \text{ and} \end{aligned} \tag{11}$$

$$0 = \sup_x\{h(x)-h(\bar{x})-\langle x^*,x-\bar{x}\rangle\colon x\epsilon F\} = -h(\bar{x})+\langle x^*,\bar{x}\rangle-h_*(x^*) \tag{12}$$

The result follows from Equations (11) and (12).

Figure 3 shows an illustration of Theorem 10 above. The primal problem to $\mathrm{minimize}_x\{g(x)-h(x)\colon x\epsilon E\cap F\}$ has an optimal solution $\bar{x}\epsilon E\cap F$. The figure shows the separating hyperplane H between A and B at $(\bar{x},h(\bar{x}))$.

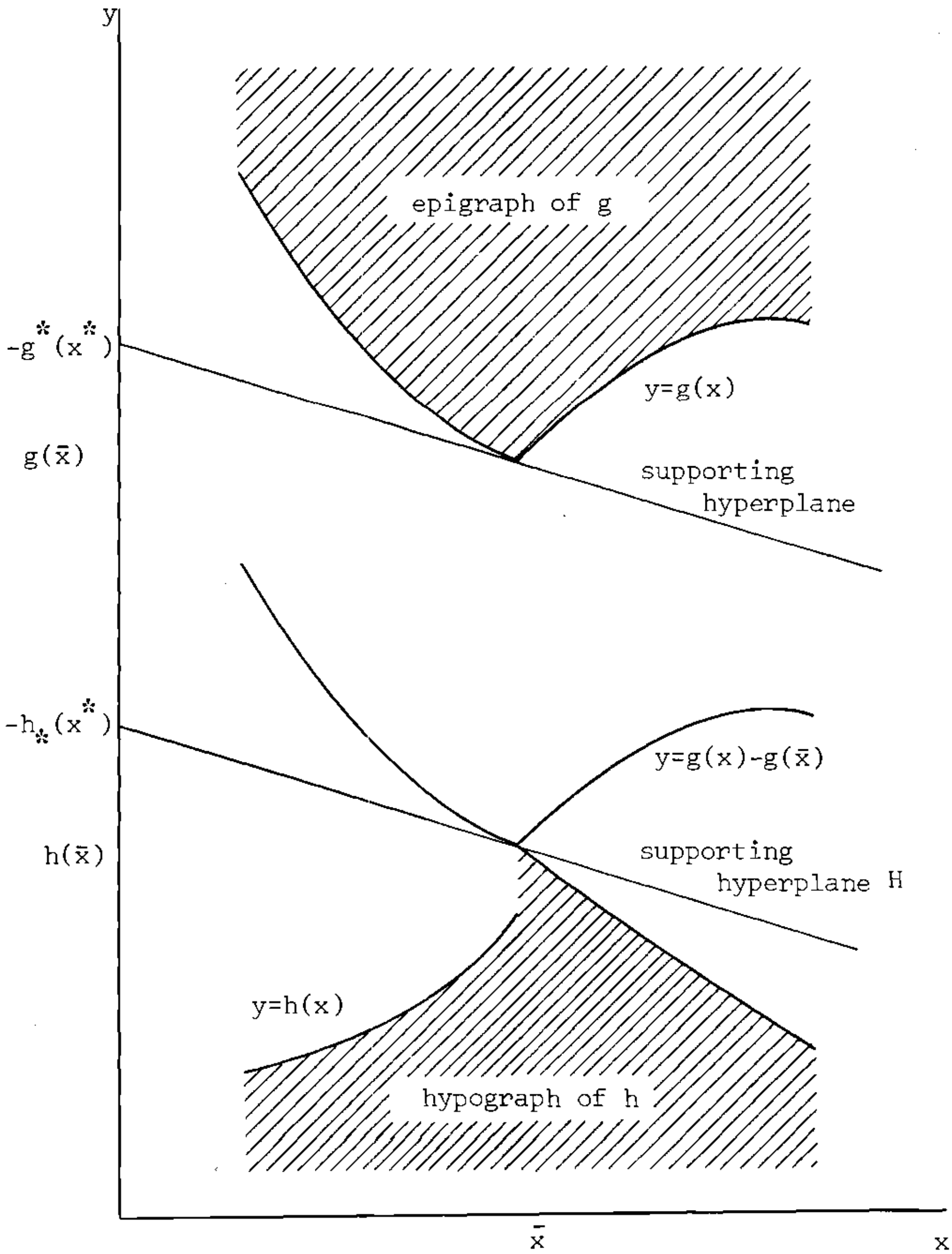

Figure 3. An Illustration of Fenchel's Duality Theorem

The intercept of the hyperplane H on the y-axis is $-h_*(x^*)$ and the intercept of the parallel hyperplane that supports the epigraph of g at $(\bar{x},g(\bar{x}))$, on the y-axis is $-g^*(x^*)$. The figure illustrates that $g(\bar{x})-h(\bar{x}) = h_*(x^*)-g^*(x^*)$, where $(x^*,-1)$ corresponds to the hyperplane H. It should be noted that x^* is a subgradient of g and h at $\bar{x}$.

<u>Nonlinear Programming via Conjugate Functions</u>

At this stage it is worthwhile discussing how a nonlinear programming problem can be expressed in the form, $\text{minimize}_x\{g(x)-h(x): x \epsilon E \cap F\}$. This can be done in two ways as follows. First we let $g(x) = f_o(x)$ for all $x \epsilon E = E^n$, and $h(x) = 0$ for all $x \epsilon F = \{x: f_i(x) \leq 0, i=1,2,\ldots,m\}$ where the functions $f_o,f_1,f_2,\ldots,f_m$ correspond to the nonlinear program, $\text{minimize}_x\{f_o(x): f_i(x) \leq 0, i=1,2,\ldots,m\}$. The problem D of Equation (6) becomes to $\text{maximize}_{x^*}\{h_*(x^*)-g^*(x^*): x^* \epsilon E^* \cap F^*\}$, where

$$g^*(x^*) = \sup_x\{<x^*,x>-f_o(x): x \epsilon E^n\} \quad \text{for all } x^* \epsilon E^*, \quad \text{where}$$

$$E^* = \{x^*: \sup_x\{<x^*,x> - f_o(x): x \epsilon E^n\} < \infty\}$$

and

$$h_*(x^*) = \inf_x\{<x^*,x>: f_i(x) \leq 0, i=1,2,\ldots,m\}$$

for all $x^* \epsilon F^* = \{x^*: \inf_x\{<x^*,x>: f_i(x) \leq 0, i=1,2,\ldots,m\} > \infty\}$.

If the functions f_o and f_i are convex, then the evaluation of g^* at an x^* is equivalent to maximizing an unconstrained concave function, whereas the evaluation of h_* at an x^* involves the minimization of a linear function over a convex set. Whether problem D above is in fact easier or more difficult to solve than the original problem, depends on the nature of the problem under consideration.

Another way to formulate a nonlinear programming problem using conjugate functions is as follows: the problem is changed into an unconstrained problem by introducing a perturbation vector, the dual of which is the price vector (lagrangian multiplier vector). See, for example, Williams [54] and Rockafeller [44]. This motivates the duality formulation via the lagrangian multiplier vector discussed in detail in Section 3. Consider the problem, $\text{minimize}_x\{f_o(x): f_i(x) \geq b_i, i=1,2,...,m, x \geq 0\}$. It is well known that this constrained problem with the n dimensional decision variable x can be transformed into an equivalent problem with an m dimensional decision vector u, usually referred to as the perturbation vector. Remark 7 below gives the form of the unconstrained problem and shows the equivalence referred to above.

Remark 7

The two problems,

$$\text{minimize}_x\{f_o(x): x \geq 0,\ f_i(x) \geq b_i,\ i=1,2,...,m\}, \text{ and}$$

$$\text{minimize}_u\{g(u) - h(u): u\epsilon E^m\} \text{ where,}$$

$$g(u) = \inf_x\{f_o(x): x \geq 0, f_i(x) \geq u_i, i=1,2,\ldots,m\} \qquad u\epsilon E^m$$

$$h(u) = \begin{cases} 0 & \text{if } u_i \geq b_i \text{ for all } i\epsilon\{1,2,\ldots,m\} \\ -\infty & \text{if } u_i < b_i \text{ for some } i\epsilon\{1,2,\ldots,m\}. \end{cases}$$

are equivalent, in the sense that,

(i) If $\bar{x}$ solves the first problem, then $\bar{u} = f(\bar{x}) = (f_1(\bar{x}),\ldots, f_m(\bar{x}))$ solves the second problem.

(ii) If $\bar{u}$ solves the second problem and the inf is attained at $\bar{x}$, then $\bar{x}$ solves the first problem.

(iii) If $\bar{x}$ solves the first problem and $\bar{u}$ solves the second problem, then $f_o(\bar{x}) = g(\bar{u}) - h(\bar{u})$.

Proof. Let $D(u) = \{x: x\geq 0, f_i(x)\geq u_i, i=1,2,\ldots,m\}$. Therefore, if $u\geq b$, then $D(u) \subset D(b)$. This implies that $g(b) \leq g(u)$ for all $u\geq b$. Since $\bar{x}$ solves the first problem, then $g(f(\bar{x})) = g(b) \leq g(u)$ for all $u \geq b$. But since $h(u) = -\infty$ if $u_i < b_i$ for some i, then $g(f(\bar{x})) \leq g(u) - h(u)$ for all $u\epsilon E^m$, which proves assertion (i). If $\bar{u}\epsilon E^m$ solves the second problem, then $\bar{u} \geq b$, and hence $g(\bar{u}) - h(\bar{u}) = g(\bar{u}) \leq g(b)$. But since $g(b) \leq g(\bar{u})$, then $g(\bar{u}) = g(b)$. But since the inf is attained at $\bar{x}$, then $\bar{x}$ solves the first problem. To prove (iii) assume that $\bar{x}$ solves the first problem and $\bar{u}$ solves the second problem. Then $g(\bar{u}) - h(\bar{u}) \leq g(b)$, but on the other hand $\bar{u} \geq b$, and hence $g(\bar{u}) - h(\bar{u}) = g(\bar{u}) \geq g(b)$, and hence the assertion is immediate.

We now consider the following two problems.

$$P:\ \text{minimize}_u\{g(u)-h(u)\} \tag{11}$$

$$D:\ \text{maximize}_{u^*}\{h_*(u^*)-g^*(u^*)\} \tag{12}$$

The following remark gives a useful form of problem D above, which is then used to give an interesting economic interpretation due to Williams [54]. It may be noticed that the vector u^* is precisely the lagrangian multiplier vector (price vector) as seen later in the economic interpretation.

Remark 8

The two problems,

$$\text{minimize}_x\{f_o(x):\ x\geq 0,\ f_i(x)\geq b_i,\ i=1,2,\ldots,m\},\ \text{and}$$

$$\text{maximize}_{u^*}\{\langle u^*,b\rangle - \sup_x\{\langle u^*,f(x)\rangle - f_o(x):\ x \geq 0\}:\ u^* \geq 0\}$$

are subdual programs.

Proof. We need consider the problem, $\text{minimize}_u\{g(u) - h(u): u\epsilon E^m\}$. We first construct g^* and h_* keeping in mind that $E = F = E^m$.

$$g^*(u^*) = \sup_u\{\langle u^*,u\rangle - g(u):\ u\epsilon E^m\}$$

$$= \sup_u\{\langle u^*,u\rangle - \inf_x\{f_o(x):\ f(x)\geq u,\ x\geq 0\}:\ u\epsilon E^m\}$$

$$= \sup_u\{\langle u^*,u\rangle - \inf_{x,s}\{f_o(x): f(x)-s = u,\ x\geq 0,\ s\geq 0\}: u\in E^m\}^\dagger$$

$$= \sup_u\{-\inf_{x,s}\{f_o(x) - \langle u^*,u\rangle: f(x)-s = u,\ x\geq 0,\ s\geq 0\}: u\in E^m\}$$

$$= \sup_u\{\sup_{x,s}\{\langle u^*,u\rangle - f_o(x): f(x)-s = u,\ x\geq 0,\ s\geq 0\}: u\in E^m\}$$

$$= \sup_s\{\sup_x\{\langle u^*,f(x)\rangle - f_o(x) - \langle u^*,s\rangle: x\geq 0\}: s\geq 0\}$$

$$= \sup_x\{\sup_s\{\langle u^*,f(x)\rangle - f_o(x) - \langle u^*,s\rangle: s\geq 0\}: x\geq 0\}$$

$$= \begin{cases} \infty & \text{if } u_i^*<0 \text{ for some } i\in\{1,2,\ldots,m\} \\ \sup_x\{\langle u^*,f(x)\rangle - f_o(x): x\geq 0\} & \text{if } u^*\geq 0 \end{cases} \qquad (13)$$

We now consider h_*.

$$h_*(u^*) = \inf_u\{\langle u^*,u\rangle - h(u): u\in E^m\}$$

$$= \inf_u\{\langle u^*,u\rangle: u\geq b\}$$

$$= \begin{cases} -\infty & \text{if } u_i^*<0 \text{ for some } i\in\{1,2,\ldots,m\} \\ \langle u^*,b\rangle & \text{if } u^*\geq 0 \end{cases} \qquad (14)$$

From this it is immediate that $E^*=F^*=E_+^m$ and hence $E^*\cap F^*=E_+^m$. Therefore

---

†At this step we introduced the slack vector $s\geq 0$ for convenience. We change the inequality $f(x)\geq u$ to the equality $f(x) - s = u$, and then eliminate the vector u by substitution.

from (13) and (14) and by Remark 6, it follows that $\min_u\{g(u) - h(u): u\epsilon E^m\} \geq \max_{u^*}\{<u^*,b> - \sup_x\{<u^*,f(x)> - f_o(x): x\geq 0\}: u^*\geq 0\}$. But by Remark 7, then it follows immediately that $\min_x\{f_o(x): f(x) \geq b, x \geq 0\} \geq \max_{u^*}\{<u^*,b> - \sup_x\{<u^*,f(x)> - f_o(x): x \geq 0\}: u^* \geq 0\}$. This completes the proof.

The above relationship has an interesting economic interpretation due to Williams [54]. Consider a manufacturer who wants to produce commodities 1,2,..., and m the demands of which are $b_1,b_2,\ldots,$ and b_m. Let the input decision vector be x, which may represent raw materials, manpower, machine hours, etc. required to do the job. Hence we have the nonnegativity constraint $x\geq 0$. Given that an input vector x is employed, then the number of units of commodity i which is produced is given by $f_i(x)$. Therefore the demand constraint becomes $f(x) \geq b$. If the cost at a level x is given by $f_o(x)$, then the manufacturer's problem is precisely the original problem, i.e., $\text{minimize}_x\{f_o(x): x\geq 0, f(x)\geq b\}$. We will now discuss the interpretation of the second problem of Remark 8. Consider a contractor who wants to rent the facilities from the manufacturer, produce the commodities, and then sell them back to him. Now the manufacturer may agree to let the contractor use his facilities if the rent paid to him by the contractor is at least equal to his maximum profit had he undertaken the whole operation by himself. In other words, suppose that the contractor quotes a price $u_i^* \geq 0$ for commodity i, then had he produced the commodities, the manufacturer could have achieved a maximum profit which is equal to $\sup_x\{<u^*,f(x)> - f_o(x): x\geq 0\}$. This profit is exactly the amount of rent the manufacturer should ask for. Therefore, from the contractor's point of view, the problem is reduced to quoting the optimal price vector u^* which maximizes his profit, namely the income $\langle u^*,b\rangle$ minus the rent, i.e. $\sup_x\{\langle u^*,f(x)\rangle - f_o(x): x\geq 0\}$. Therefore the contractor's problem becomes to $\text{maximize}_{u^*}\{\langle u^*,b\rangle - \sup_x\{\langle u^*,f(x)\rangle - f_o(x): x\geq 0\}: u^*\geq 0\}$, or precisely problem D above. Hence the relationship derived earlier states that the manufacturer's minimum cost is larger than or equal to the contractor's maximum profit. Moreover, if the functions involved satisfy the hypotheses of Theorem 6, then the two optimal solutions are equivalent.

3. Duality and the Lagrangian Multiplier Vector

The approach used above has lead to the introduction of a perturbation vector u and a lagrangian multiplier vector u^*. This motivates considering a different form of duality formulation via the lagrangian multiplier vector. We present two problems, due to Falk [17], and show that they are subdual programs. Under certain convexity assumptions, Falk showed that the two problems are indeed dual programs. We show that the two problems can be derived from the Minmax formulation. This permits us to apply Theorems 1 through 4, which give results about the equivalence of optimal solutions and existence of these optimal solutions. These results are extensions of Falk's duality theorem. We then investigate the relationship between this formulation and the conjugate function formulation.

Falk's Theorem and Extensions

Falk [17] considered problem P, to $\text{minimize}_x\{f_o(x)\colon x\epsilon\Omega,$ $f(x)\le 0\}$ where $\Omega\subset E^n$ and $f(x) = (f_1(x),\ldots,f_m(x))$ for any $x\epsilon E^n$. In order to formulate problem D, which is referred to by Falk as the auxiliary problem we introduce the following two functions ϕ and β. Let $\phi(x,u^*) = f_o(x) + \langle u^*,f(x)\rangle$ for all $x\epsilon E^n$ and $u^*\epsilon E^m$. We now consider the induced function $\beta\colon E^m\to E^1$ defined by,

$$\beta(u^*) = \inf_x\{\phi(x,u^*)\colon x\epsilon\Omega\} \qquad \text{for all } u^*\epsilon U^* \text{ where,}$$

$$U^* = \{u^*\colon u^*\epsilon E^m,\ \inf_x\{\phi(x,u^*)\colon x\epsilon\Omega\} > -\infty\}.$$

In this section, problems P and D are given by Equations (15) and (16) below.

$$\text{P: minimize}_x\{f_o(x)\colon x\epsilon\Omega,\ f(x)\le 0\} \tag{15}$$

$$\text{D: maximize}_{u^*}\{\beta(u^*)\colon u^*\ge 0,\ u^*\epsilon U^*\}. \tag{16}$$

The following remark shows that U^* is a convex set and that the function β is concave. It should be noted that no convexity or concavity assumption of f_o or f is required.

Remark 9

The function $\beta\colon E^m\to E^1$ where $\beta(u^*) = \inf_x\{f_o(x) + \langle u^*,f(x)\rangle\colon$ $x\epsilon\Omega\}$ is concave and the set $U^* = \{u^*\colon \inf_x\{\phi(x,u^*)\colon x\epsilon\Omega\} > -\infty\}$ is convex. The proof is similar to that of Remark 1 and is omitted.

The following remark is a slight extension of a result by Falk [17] where the differentiability assumption of β is relaxed. The proof is the same, however. We also include a result of Everett [16] usually referred to as the generalized lagrangian multiplier theorem. The connection was given by Falk [17].

Remark 10

If $\bar{x}\epsilon\Omega$ solves the problem, $\text{minimize}_x\{f_o(x) + \langle\bar{u}^*,f(x)\rangle: x\epsilon\Omega\}$ for some $0 \leq \bar{u}^*\epsilon U^*$, then $f(\bar{x})$ is a subgradient of β at $\bar{u}^*$. Moreover $\bar{x}$ solves the problem, $\text{minimize}_x\{f_o(x): x\epsilon\Omega,\ f(x) \leq \{f(\bar{x})\}$.

Proof. $\beta(u^*) = \inf_x\{f_o(x) + \langle u^*,f(x)\rangle: x\epsilon\Omega\} \leq f_o(\bar{x}) + \langle u^*,f(\bar{x})\rangle$

$$= f_o(\bar{x}) + \langle\bar{u}^*,f(\bar{x})\rangle - \langle(\bar{u}^*,f(\bar{x})\rangle + \langle u^*,f(\bar{x})\rangle$$

$$= \inf_x\{f_o(x) + \langle\bar{u}^*,f(x)\rangle: x\epsilon\Omega\} + \langle f(\bar{x}),u^*-\bar{u}^*\rangle$$

$$= \beta(\bar{u}^*) + \langle f(\bar{x}),u^*-\bar{u}^*\rangle$$

The above inequality that shows that $f(\bar{x})$ is a subgradient of β at $\bar{u}^*$. On the other hand, by hypothesis, $f_o(\bar{x}) \leq f_o(x) + \langle\bar{u}^*,f(x) - f(\bar{x})\rangle$ for all $x\epsilon\Omega$. But since $\bar{u}^* \geq 0$, then $f_o(\bar{x}) \leq f_o(x)$ if $f(x) \leq f(\bar{x})$. This implies that $\bar{x}$ solves the problem, $\text{minimize}_x\{f_o(x): x\epsilon\Omega,$ $f(x) \leq f(\bar{x})\}$.

We now turn to the problems P and D of Equations (15) and (16) and show that they are subdual programs.

Remark 11

The two problems P and D are subdual programs.

Proof. We show that $f_o(\bar{x}) \geq \beta(\bar{u}^*)$ for an arbitrary feasible solutions $\bar{x}$ and $\bar{u}^*$ of problems P and D. Noting that $f(\bar{x}) \leq 0$ and $\bar{u}^* \geq 0$ we conclude that, $\beta(\bar{u}^*) = \inf_x\{f_o(x) + \langle\bar{u}^*,f(x)\rangle: x\epsilon\Omega\} \leq f_o(\bar{x}) + \langle\bar{u}^*,f(\bar{x})\rangle \leq f_o(\bar{x})$. This completes the proof.

By letting the functions $f_o,f_1,\ldots,$ and f_m be convex and the set Ω be convex, the equivalence between the solutions of the two problems is established. This result is given by the following theorem. For a proof refer to [17] and [32].

Theorem 11

Let $f_i: E^n \to E^1$ $(i=0,1,2,\ldots,m)$ be convex functions and let $f(x) = (f_1(x),f_2(x),\ldots,f_m(x))$ for all $x\epsilon E^n$. Let $\Omega \subset E^n$ be a convex set, and consider the problem P, to minimize$_x\{f_o(x): f(x) \leq 0, x\epsilon\Omega\}$. Assume that there exists some $x\epsilon\Omega$ such that $f(x) < 0$. Let $\mu = \inf_x\{f_o(x): f(x) \leq 0, x\epsilon\Omega\}$ be finite and let $\beta(u^*) = \inf_x\{f_o(x) + \langle u^*,f(x)\rangle: x\epsilon\Omega\}$ for all $u^* \geq 0$ and $u^*\epsilon U^* = \{u^*: u^*\epsilon E^m, \inf_x\{f_o(x) + \langle u^*,f(x)\rangle: x\epsilon\Omega\} > -\infty\}$. Then there exists $0 \leq \bar{u}^*\epsilon U^*$ such that $\mu = \max_{u^*}\{\beta(u^*): u^*\geq 0, u^*\epsilon U^*\} = \beta(\bar{u}^*)$. Moreover, if μ is achieved by some $\bar{x}\epsilon\Omega$, and $f(\bar{x}) \leq 0$, then $\langle\bar{u}^*,f(\bar{x})\rangle = 0$, and $\bar{x}$ solves the problem: minimize$_x\{f_o(x) + \langle\bar{u}^*,f(x)\rangle: x\epsilon\Omega\}$.

We now show that Falk's formulation of duality is a special case of the Minmax formulation. We let ϕ: $E \times F \to E^1$ be defined as follows. $\phi(x,u^*) = f_o(x) + \langle u^*,f(x)\rangle$ for all $x\epsilon E$, and all $u^*\epsilon F$. Further, we let $E = \Omega$, and $F = E^m_+$. In order to formulate problems P and D of Equations

(3) and (4), we develop the functions α and β as follows,

$$\alpha(x) = \sup_{u^*}\{\phi(x,u^*)\colon u^* \geq 0\} = \sup_{u^*}\{f_o(x) + \langle u^*, f(x)\rangle\colon u^* \geq 0\} \quad \text{for all } x \in \Omega$$

$$= \infty \qquad \text{if } f_i(x) > 0 \text{ for some } i \in \{1,2,\ldots,m\}$$

$$f_o(x) \qquad \text{if } f(x) \leq 0 \text{ for } x \in \Omega$$

$$\beta(u^*) = \inf_x\{\phi(x,u^*)\colon x \in \Omega\} = \inf_x\{f_o(x) + \langle u^*, f(x)\rangle\colon x \in \Omega\} \quad \text{for } u^* \geq 0.$$

It then follows that the problems P and D are identical to the two problems of Falk. Therefore, we can apply the results of the Min-max theory of Section 1 to obtain conditions under which problems P and D become dual programs and also to obtain some theorems of existence of optimal solutions to both problems. This leads to Theorem 12 through 15 which give different versions, as well as extensions of Falk's results.

Theorem 12

Let $\phi(x,u^*) = f_o(x) + \langle u^*, f(x)\rangle$ for all $x \in \Omega$ and all $u^* \geq 0$, where Ω is a closed convex set in E^n, and $f(x) = (f_1(x),\ldots,f_m(x))$ for all $x \in E^n$. Let $f_o, f_1, \ldots,$ and f_m be *l.s.c.* and let $\phi(.,u^*)$ be quasi-convex on Ω for each $u^* \geq 0$. Then by Theorem 1,

$$\inf_x\{f_o(x)\colon f(x) \leq 0,\ x \in \Omega\} = \sup_{u^*}\{\inf_x\{f_o(x) + \langle u^*, f(x)\rangle\colon x \in \Omega\}\colon u^* \geq 0\}.$$

Theorem 13

Let $\phi(x,u^*) = f_o(x) + \langle u^*,f(x)\rangle$ for all $u^*\epsilon F = E_+^m$ and $x\epsilon\Omega$, where Ω is a closed convex set in E^n, and $f(x) = (f_1(x),\ldots,f_m(x))$ for all $x\epsilon E^n$. Let $\beta(u^*) = \inf_x\{\phi(x,u^*)$: $x\epsilon\Omega\}$ for all $u^* \geq 0$. Suppose that $f_o,f_1,\ldots,f_m$ are convex and *l.s.c.* on Ω. Then by Theorem 2 if condition (i) below is satisfied, then,

$$\max_{u^*}\{\beta(u^*): u^*\geq 0\} = \inf_x\{f_o(x): f(x)\leq 0,\ x\epsilon\Omega\} < \infty.$$

If condition (ii) below is satisfied, then by Theorem 2,

$$\sup_{u^*}\{\beta(u^*): u^*\geq 0\} = \min_x\{f_o(x): f(x)\leq 0,\ x\epsilon\Omega\} > -\infty.$$

If both conditions are satisfied, then

$$\max_{u^*}\{\beta(u^*): u^*\geq 0\} = \min_x\{f_o(x): f(x)\leq 0,\ x\epsilon\Omega\}.$$

Condition (i). No nonzero vector $\overset{o}{u}{}^*$ has the property that all $u^*\epsilon r(F)$ and $x\epsilon r(\Omega)$ the ray $\{u^* + \lambda\overset{o}{u}{}^*: \lambda\geq 0\}$ is contained in F, and $\phi(x,u^*+\lambda\overset{o}{u}{}^*)$ is a nonzero decreasing function of $\lambda \geq 0$.

Condition (ii). No nonzero vector x^o has the property that for all $u^*\epsilon r(F)$ and $x\epsilon r(\Omega)$ the ray $\{x+\lambda x^o: \lambda\geq 0\}$ is contained in Ω, and $\phi(x+\lambda x^o,u^*)$ is a nonincreasing function of $\lambda\geq 0$.

Theorem 14

Let $\phi(x,u^*) = f_o(x) + \langle u^*,f(x)\rangle$ for all $u^*\epsilon F = E_+^m$ and all $x\epsilon\Omega$, where Ω is a closed and convex set in E^n and $f(x) = (f_1(x),\ldots,f_m(x))$ for all $x\epsilon E^n$. Let $\beta(u^*) = \inf_x\{\phi(x,u^*): x\epsilon\Omega\}$ for all $u^* \geq 0$. Suppose that $f_o,f_1,\ldots,$ and f_m are continuous and convex on Ω, and that $\bar{x}$ solves problem P as defined by Equation (15) and assume that $\bar{u}^*\epsilon F$ is associated with $\bar{x}$. Then by Theorem 3, there exists $\overset{o}{u}{}^*\epsilon E^m$ which solves problem D as defined by Equation (16), and $(\bar{x},\overset{o}{u}{}^*)$ is a saddle point of ϕ on ΩxF iff ϕ has the low value property at $(\bar{x},\bar{u}^*)$. Further suppose that $\hat{u}^*$ solves problem D and that $\hat{x}\epsilon\Omega$ is associated with $\hat{u}^*$. Then by Theorem 3 there exists an x^o that solves problem P and $(x^o,\hat{u}^*)$ is a saddle point of ϕ on ΩxF iff ϕ has the high value property at $(\hat{x},\hat{u}^*)$.

Theorem 15

Let $\phi(x,u^*) = f_o(x) + \langle u^*,f(x)\rangle$ for all $u^*\epsilon F = E_+^m$ and all $x\epsilon\Omega$, where Ω is a closed and convex set in E^n and $f(x) = (f_1(x),\ldots,f_m(x))$ for all $x\epsilon E^n$. Let $\beta(u^*) = \inf_x\{\phi(x,u^*): x\epsilon\Omega\}$ for all $u^* \geq 0$. Assume that $f_o,f_1,\ldots,$ and f_m are *l.s.c.* and that $\phi(.,u^*)$ is *s*-strictly quasi-convex on Ω for each $u^*\geq 0$. Let $\bar{x}$ solve problem P as defined by Equation (15) and assume that $\bar{u}^*\epsilon F$ is associated with $\bar{x}$. If $\phi(\bar{x},.)$ is strictly quasi-concave in some neighborhood of $\bar{u}^*$, then by Theorem 4, $\bar{u}^*$ solves problem D as defined by Equation (16) and $(\bar{x},\bar{u}^*)$ is a saddle point of ϕ on ΩxF. Further let $\overset{o}{u}{}^*$ solve problem D and assume that $\overset{o}{x}\epsilon\Omega$ is associated with $\overset{o}{u}{}^*$. If $\phi(.,\overset{o}{u}{}^*)$ is strictly quasi-convex in some neighborhood of x^o, then by Theorem 4, x^o solves problem P and $(x^o,\overset{o}{u}{}^*)$ is a saddle point of ϕ on ΩxF.

<u>The Relationship between the Conjugate and the Lagrangian Multiplier Formulations</u>

It may be mentioned that in general the conjugate function and the lagrange multiplier formulations of duality are not equivalent. However, we show below that by confining the problem P to $\text{minimize}_x\{f_o(x)\text{: } x\geq 0,\ f_i(x)\leq 0,\ i=1,2,\ldots,m\}$, the problems D according to Fenchel and Falk are equivalent (recall that problems P and D are subdual programs). This is done via two different routes. First by introducing an n dimensional vector x^* and secondly by introducing an m dimensional vector u^*.

We first show the equivalence by introducing the "lagrangian" vector x^*. To formulate problem D due to Falk, which corresponds to Equation (16), consider the following.

Let $\phi\text{:}E^n x E^n \to E^1$, where $\phi(x,x^*) = f_o(x) - \langle x^*,x\rangle$, $\Omega = \{x\text{: } x\epsilon E^n,\ f(x)\leq 0\}$, and $\beta\text{: } E^n \to E^1$, where $\beta(x^*) = \inf_x\{f_o(x) - \langle x^*,x\rangle\text{: } f(x)\leq 0\}$ for all $x^*\geq 0$, $x^*\epsilon X^*$ where $X^* = \{x^*\text{: } \inf_x\{f_o(x) - \langle x^*,x\rangle\text{: } f(x)\leq 0\} > -\infty\}$.

Therefore, problem D corresponding to Equation (16) becomes to $\text{maximize}_{x^*}\{\inf_x\{f_o(x) - \langle x^*,x\rangle\text{: } f(x)\leq 0\}\text{: } x^*\geq 0,\ x^*\epsilon X^*\}$.

On the other hand, in the conjugate function formulation, the problem P is of the form, $\text{minimize}_x\{g(x) - h(x)\text{: } x\epsilon E\cap F\}$. So we consider $g\text{: } E\to E^1$ and $h\text{: } F\to E^1$. Let $g = f_o$, $h = 0$, $E = \{x\text{: } f(x)\leq 0\}$, and $F = E^n_+$. Therefore,

$$g^*(x^*) = \sup_x\{<x^*,x> - f_o(x): f(x)\leq 0\} \text{ for all } x^*\epsilon E^*, \text{where}$$

$$E^* = \{x^*: \sup_x\{<x^*,x> - f_o(x): f(x)\leq 0\} < \infty\}$$

$$= \{x^*: \inf_x\{f_o(x) - <x^*,x>: f(x)\leq 0\}>-\infty\} = X^*$$

$$h_*(x^*) = \inf_x\{<x^*,x>: x\geq 0\} = \begin{cases} -\infty & \text{if } x_i^*<0 \text{ for some } i\epsilon\{1,2,\ldots,m\} \\ 0 & \text{if } x^*\geq 0. \end{cases}$$

Therefore $F^* = E_+^n$, and hence problem D according to Fenchel is to maximize$_{x^*}\{-\sup_x\{<x^*,x> - f_o(x): f(x)\leq 0 : x^*\epsilon X^*, x^*\geq 0\}$, or in other words, maximize$_{x^*}\{\inf_x\{f_o(x) - <x^*,x>: f(x)\leq 0 : x^*\epsilon X^*, x^*\geq 0\}$. This is precisely Falk's problem obtained above.

We can also show the equivalence by introducing an m dimensional lagrangian vector u^*. We first consider the lagrangian formulation. To develop problem D defined by Equation (16), consider the following.

Let ϕ: $E^n x E^m \to E^1$ be given by $\phi(x,u^*) = f_o(x) + <u^*,f(x)>$, and let $\Omega = E_+^n$. Consider β: $E^m \to E^1$, where $\beta(u^*) = \inf_x\{f_o(x) + <u^*,f(x)>: x\geq 0\}$ for all $u^* \geq 0$, and $u^*\epsilon U^* = \{u^*: \inf_x\{f_o(x) + <u^*,f(x)>: x\geq 0\} > -\infty\}$. Therefore problem D becomes to maximize$_{u^*}\{\inf_x\{f_o(x) + <u^*,f(x)>: x\geq 0\}: u^*\geq 0, u^*\epsilon U^*\}$.

Now we consider problem D via conjugate functions. For this purpose we change the problem: minimize$_x\{f_o(x): f(x)\leq 0, x\geq 0\}$ to an equivalent unconstrained problem of the form, minimize$_u\{g(u)-h(u): u\epsilon E^m\}$, where $g(u) = \inf_x\{f_o(x): f(x)\leq u, x\geq 0\}$ for all $u\epsilon E^m$, and where $h(u) = 0$

if $u \leq 0$ and $h(u) = -\infty$ if $u_i > 0$ for some $i \in \{1,2,\ldots,m\}$. An argument similar to that used in proving Remark 7 can be used to show the equivalence between the two problems. Hence to construct problem D, we develop the functions g^* and h_* as follows.

$$
\begin{aligned}
g^*(u^*) &= \sup_u \{<u^*,u> - g(u) : u \in E^m\} \\
&= \sup_u \{<u^*,u> - \inf_x \{f_o(x) : f(x) \leq u,\ x \geq 0\} : u \in E^m\} \\
&= \sup_u \{<u^*,u> - \inf_{x,s} \{f_o(x) : f(x)+s = u,\ s \geq 0,\ u \geq 0\} : u \in E^m\} \\
&= \sup_u \{\sup_{x,s} \{<u^*,u> - f_o(x) : f(x) + s = u,\ s \geq 0,\ u \geq 0\} : u \in E^m\} \\
&= \sup_x \{\sup_s \{<u^*,f(x)> + <u^*,s> - f_o(x) : s \geq 0\} : x \geq 0\} \\
&= \infty \qquad \text{if } u_i^* > 0 \text{ for some } i \in \{1,2,\ldots,m\} \qquad (17) \\
&\quad\ \sup_x \{<u^*,f(x)> - f_o(x) : x \geq 0\} \qquad \text{if } u^* \leq 0.
\end{aligned}
$$

$$
\begin{aligned}
h_*(u^*) &= \inf_u \{<u^*,u> - h(u) : u \in E^m\} \\
&= \inf_u \{<u^*,u> : u \leq 0\} \\
&= -\infty \qquad \text{if } u_i^* > 0 \text{ for some } i \in \{1,2,\ldots,m\} \qquad (18) \\
&\quad\ 0 \qquad \text{if } u^* \leq 0.
\end{aligned}
$$

From (17) and (18), the problem becomes,

$\text{maximize}_{u^*}\{-\sup_x\{<u^*,f(x)> - f_o(x): x \geq 0\}: u^* \leq 0\}$, which is equivalent to, $\text{maximize}_{u^*}\{\inf_x\{f_o(x) - <u^*,f(x)>: x \geq 0 : u^* \leq 0\}$. But this can be written as, $\text{maximize}_{u^*}\{\inf_x\{f_o(x) + <u^*,f(x): x \geq 0\}: u^* \geq 0, u^* \epsilon U^*\}$, which is the same problem obtained earlier via Falk's formulation. This again shows the equivalence between problems D according to Fenchel and Falk, given that the original problem is to $\text{minimize}_x\{f_o(x): f(x) \leq 0, x \geq 0\}$.

In the above discussion, while considering Fenchel's formulation, we introduced the function g defined by $g(u) = \inf_x\{f_o(x): f_i(x) \leq u_i$, $i=1,2,\ldots,m, x \geq 0\}$, where u is an m dimensional perturbation vector. To relate the function g to the lagrangian multiplier formulation, the nonnegativity constraints can be replaced by the more general constraint set $\Omega \subset E^n$. It is obvious that the value of the perturbation function g at $u = 0$ gives the optimal solution of the original problem. Also the g function may help in the study of a family of problems where the constraints are partially relaxed from $f(x) \leq 0$ to $f(x) \leq u$. Furthermore, the sensitivity of g to slight changes in u near the zero vector may be of interest. For an excellent discussion, see Rockafeller [44]. The following remark establishes the relationship between the conjugate of g and the β function referred to earlier while discussing Fenchel's formulation. Luenberger [32] stated that there is a relationship between g^* and β.

<u>Remark 12</u>

Let $g: E^m \to E^1$ be defined by $g(u) = \inf_x\{f_o(x): x \epsilon \Omega, f(x) \leq u\}$ where $\Omega \subset E^n$, $f_o, f_1, \ldots,$ and, $f_m: E^n \to E^1$ and $f(x) = (f_1(x), \ldots, f_m(x))$ for all $x \epsilon E^n$. Then the conjugate of g corresponds to $\beta(u^*) = \inf_x\{f_o(x) +$

$\langle u^*, f(x)\rangle$: $x\in\Omega\}$ for $u^* \geq 0$.

Proof. $g^*(u^*) = \sup_u\{\langle u^*,u\rangle - \inf_x\{f_o(x)\colon f(x)\le u,\ x\in\Omega\}\colon u\in E^m\}$

$$= \sup_u\{\langle u^*,u\rangle - \inf_{x,s}\{f_o(x)\colon$$

$$f(x) + s = u,\ x\in\Omega,\ s\ge 0\}\colon u\in E^m\}$$

$$= \sup_u\{\sup_{x,s}\{\langle u^*,u\rangle - f_o(x)\colon$$

$$f(x) + s = u,\ x\in\Omega,\ s\ge 0\}\colon u\in E^m\}$$

$$= \sup_x\{\sup_s\{\langle u^*,f(x)\rangle + \langle u^*,s\rangle - f_o(x)\colon s\ge 0\}\colon x\in\Omega\}$$

$$= \sup_x\{\langle u^*,f(x)\rangle - f_o(x)\colon x\in\Omega\} \quad \text{if } u^*\le 0 \tag{19}$$

$$\infty \qquad \text{if } u_i^*>0 \text{ for some } i\in\{1,2,\ldots,m\}$$

From (19) we conclude that, if $u^*\le 0$, then

$$g^*(u^*) = -\inf_x\{f_o(x) + \langle -u^*,f(x)\rangle\colon x\in\Omega\} = -\beta(-u^*).$$

This completes the proof.

The above result can be rewritten in the form,

$$\beta(-u^*) = -g^*(u^*) = -\sup_u\{\langle u^*,u\rangle - g(u)\colon u\in E^m\} = \inf_u\{g(u)+\langle -u^*,u\rangle\colon u\in E^m\},$$

for $u^* \leq 0$, which can further be written as,

$$\beta(u^*) = \inf_u\{g(u) + \langle u^*,u\rangle : u\epsilon E^m\} \quad \text{for } u^*\geq 0. \tag{20}$$

It should be noted that Equation (20) is equivalent to a similar result in Luenberger [32].

The geometric interpretation of the above result is illustrated in Figure 4. For a given $u^*\leq 0$, we construct the hyperplane $y = \langle u^*,u\rangle$. The distance between the surface $y = g(u)$ and the hyperplane is minimized, and hence the hyperplane is moved vertically parallel to itself until it supports the epigraph of g. The result asserts that the intercept of the hyperplane on the y-axis is equal to $\beta(-u^*)$. If the $\inf_u\{g(u) + \langle -u^*,u\rangle : u\epsilon E^m\}$ is attained at $\bar{u}\epsilon E^m$, then u^* is a subgradient of g at $\bar{u}$ and $(u^*,-1)$ is an *o.n.* to the epigraph of g at $(\bar{u},g(\bar{u}))$. This implies that if the surface $y = g(u)$ is tilted by adding the linear function $\langle -u^*,u\rangle$, then the absolute minimum of the resulting function $g(u) + \langle -u^*,u\rangle$ is achieved at $\bar{u}$.

Further if $f_o,f_1,\ldots,$ and f_m are convex functions, then it can be concluded that the function g is convex, see [32] for example. Hence, if we consider $u = 0$, then there is a supporting hyperplane of the epigraph of g at $(0,g(0))$, and hence there exists $(u^*,-1)$ which is an *o.n.* to the epigraph of g at $(0,g(0))$. This is illustrated in Figure 5. It should be noted that $g(0)$ is the optimal solution of the original problem and that $-u^*$ is the lagrangian multiplier vector.

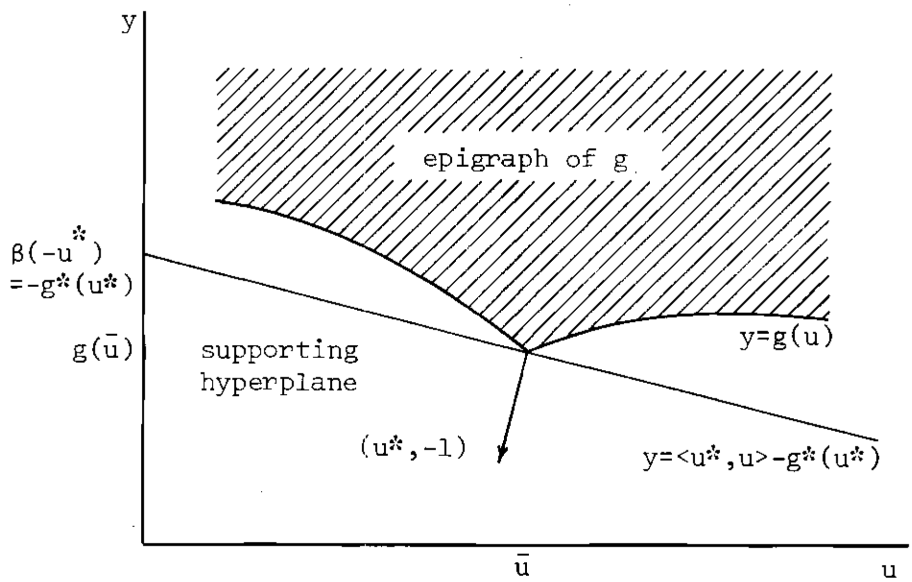

Figure 4. The Perturbation Function

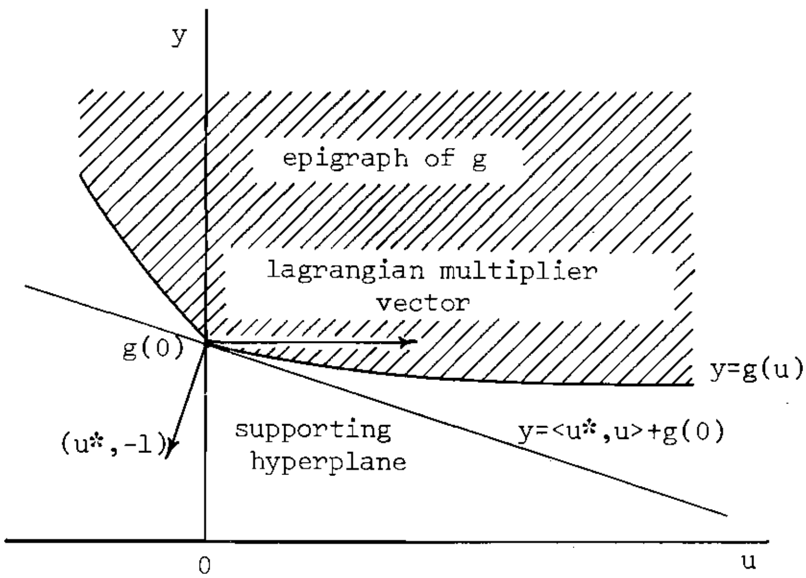

Figure 5. The Perturbation Function and the Lagrangian Multiplier Vector

In closing this section it is worthwhile mentioning that many known duality theorems and formulations are subsumed under the general formulations of the lagrangian multiplier vector and the conjugate functions, which are subsumed under the Minmax formulation, e.g. duality in linear programming, duality in quadratic programming, and Wolfe's duality theorems. For example, see [14], [26], [33], and [56].

4. Wolfe's Duality, Symmetric Duality, and Extensions

In this section we discuss Wolfe's duality formulation and the symmetric duality formulation due to Dantzig, et al. See, for example, [11], [13], [39], and [56]. We show that both can be formulated as special cases of the Minmax formulation.

Stoer [48] has showed that Wolfe's duality formulation can be obtained from Minmax formulation. We let $\phi\colon ExF \to E^1$ be defined by $\phi(x,y) = f_o(x) + \langle y, f(x)\rangle$ for all $x \in E = E^n$, and all $y \in F = E^m_+$, where $f(x) = (f_1(x), f_2(x), \ldots, f_m(x))$ for all $x \in E^n$. Problems P and D defined by Equations (3) and (4) can be established by constructing the functions α and β as follows.

$$\begin{aligned}\alpha(x) &= \sup_y\{f_o(x) + \langle y, f(x)\rangle : y \geq 0\} \\ &= f_o(x) \quad \text{if } f(x) \leq 0 \qquad (21) \\ &\quad \infty \qquad \text{if } f_i(x) > 0 \text{ for some } i \in \{1,2,\ldots,m\}.\end{aligned}$$

$$\beta(y) = \inf_x\{f_o(x) + \langle y, f(x)\rangle : x \in E^n\} \quad \text{for all } y \geq 0. \qquad (22)$$

Therefore, problems P and D become,

$$P: \text{minimize}_x\{f_o(x): f(x)\leq 0\} \tag{23}$$

$$D: \text{maximize}_y\{\inf_x\{f_o(x) + \langle y,f(x)\rangle: x\epsilon E^n\}: y\geq 0\}. \tag{24}$$

Problem P is the usual form of a nonlinear program. It has been observed by Stoer [48], Mangasarian and Ponstein [36], and Whinston [52] that problem D reduces to Wolfe's† form when the functions $f_o,f_1,\ldots,$ and f_m are convex and when the inf in Equation (22) is attained. In this case problem D becomes,

$$\text{maximize}_{x,y}\{f_o(x) + \langle y,f(x)\rangle: y\geq 0,\ \nabla f_o(x) + \sum_{i=1}^{m} y_i\nabla f_i(x) = 0\}.$$

It may be noted that the above formulation is a special case of Falk's formulation when $\Omega = E^n$. By letting Ω be E^n in Theorems 12 through 15, we obtain results concerning equivalence of the optimal solutions to problems P and D above, and also results concerning existence of solutions to both problems.

Some further results can be obtained while dealing with non-differentiable functions by replacing the gradient vectors by

---

†Notice that if the inf in Equation (22) is attained for every $y\geq 0$, then by convexity of $f_o,f_1,\ldots,$ and f_m, evaluating $\beta(y)$ is equivalent to finding an $x\epsilon E^n$ such that $\nabla f_o(x) + \sum_{i=1}^{m} y_i\nabla f_i(x) = 0$.

subgradients and by using the generalized Kuhn-Tucker conditions developed in Chapter II. To discuss this in more detail consider the following two problems.

$$P: \ \text{minimize}_x \{f_o(x): f_i(x) \le 0, \ i=1,2,\ldots,m\} \tag{25}$$

$$D: \ \text{maximize}_{x,y} \{f_o(x) + \langle y, f(x)\rangle : x \in E^n, \ 0 \le y \in E^m, \tag{26}$$

$$\theta^o(x) + \sum_{i=1}^{m} y_i \theta^i(x) = 0\}$$

where $f_i: E^n \to E^1$ $(i=0,1,2,\ldots,m)$ are convex functions, and $\theta^i(x)$ is a subgradient of f_i at x.

For convenience we denote the sets of feasible solutions of problems P and D by $X \subset E^n$ and $Y \subset E^{n+m}$, respectively.

The following remark shows that the two problems satisfy the bounding property and hence are subdual programs.

Remark 13

The problems P and D defined by Equations (25) and (26) are subdual programs.

Proof. Let $\bar{x} \in X$ and $(x,y) \in Y$. Therefore, $\theta^o(x) + \sum_{i=1}^{m} y_i \theta^i(x) = 0$. By convexity of f_o and f_i we conclude that,

$$f_o(\bar{x}) \ge f_o(x) + \langle \theta^o(x), \bar{x}-x\rangle \tag{27}$$

$$f_i(\bar{x}) \ge f_i(x) + \langle \theta^i(x), \bar{x}-x\rangle \quad \text{for all } i \in \{1,2,\ldots,m\} \tag{28}$$

By multiplying each inequality in (28) by y_i, and adding (27) and the inequalities corresponding to (28) we obtain,

$$f_o(\bar{x}) + \langle y,f(\bar{x})\rangle \geq f_o(x) + \langle y,f(x)\rangle + \langle\theta^o(x) + \sum_{i=1}^{m} y_i\theta^i(x),\bar{x}-x\rangle$$

$$= f_o(x) + \langle y,f(x)\rangle. \tag{29}$$

But since $y \geq 0$ and $f(\bar{x}) \leq 0$, then from (29), $f_o(\bar{x}) \geq f_o(x) + \langle y,f(x)\rangle$ and hence the bounding property is satisfied and the proof is complete.

The following theorem asserts that if there exists a solution to problem P above, then there exists a solution to problem D and the optimal solutions are equivalent. This establishes the fact that the two problems are dual programs.

Theorem 16

Let $\bar{x}$ solve problem P defined by Equation (25). Then there exists $\bar{y} \geq 0$ such that $(\bar{x},\bar{y})$ solve problem D as defined by Equation (26). Moreover, $\langle\bar{y},f(\bar{x})\rangle = 0$.

Proof. Let (x^1,y) and (x^2,y) be feasible solutions of problem D. By an argument similar to that used in Remark 13 above we conclude that $f_o(x^1) + \langle y,f(x^1)\rangle \geq f_o(x^2) + \langle y,f(x^2)\rangle$ and also $f_o(x^2) + \langle y,f(x^2)\rangle \geq f_o(x^1) + \langle y,f(x^1)\rangle$. This implies that if (x^1,y), $(x^2,y) \in Y$, then $\phi(x^1,y) = \phi(x^2,y)$ where $\phi(x,y) = f_o(x) + \langle y,f(x)\rangle$. It can be shown, by the saddle value theorem that there exists $\bar{y} \geq 0$ such that, $\phi(\bar{x},y) \leq \phi(\bar{x},\bar{y}) \leq \phi(x,\bar{y})$ for all $x\in E^n$, $y\geq 0$. Moreover, it can be shown that $\langle\bar{y},f(\bar{x})\rangle = 0$, and that $\theta^o(\bar{x}) + \sum_{i=1}^{m} \bar{y}_i\theta^i(\bar{x}) = 0$. See, for example, [5] and Theorem 12 of Chapter II.

Therefore, let $(x,y)\epsilon Y$, then,

$$\phi(\bar{x},\bar{y}) = \max_y\{\phi(\bar{x},y): y\geq 0\} \geq \max_y\{\phi(\bar{x},y): (\bar{x},y)\epsilon Y\}$$

$$= \max_y\{\phi(x,y): (x,y)\epsilon Y\} \tag{30}$$

Inequality (30) shows that $(\bar{x},\bar{y})$ solves problem D and the proof is complete.

Theorem 17 below is in some sense the converse of Theorem 16. The theorem asserts that if a solution of the primal problem exists, say $\bar{x}$, and there is a solution (x^o,y^o) of the dual problem, then under some strict convexity assumption, $\bar{x} = x^o$.

Theorem 17

Let $\bar{x}\epsilon E^n$ solve problem P above and $(x^o,y^o) \epsilon E^{n+m}$ solve problem D. If $\phi(.,y^o)$ is strictly convex in a neighborhood of x^o, then $\bar{x} = x^o$ and $\phi(x^o,y^o) = f_o(\bar{x})$.

Proof. Assume on the contrary that $x^o \neq \bar{x}$. By Theorem 20 above, there exists $\bar{y}\geq 0$ such that $(\bar{x},\bar{y})$ solve the dual problem. Therefore,

$$\phi(\bar{x},\bar{y}) = \phi(x^o,y^o) = \max_{x,y}\{\phi(x,y): (x,y)\epsilon Y\}$$

Since $(x^o,y^o)\epsilon Y$, then $\theta(x^o) = \theta^o(x^o) + \sum_{i=1}^{m} y_i^o\theta^i(x^o) = 0$, where $^i(x^o)$ is a subgradient of f_i at x^o, for each $i\epsilon\{0,1,2,\ldots,m\}$ and $\theta(x^o)$ is a subgradient of $\phi(.,y^o)$ at x^o). By strict convexity of $\phi(.,y^o)$ at a neighborhood of x^o, it follows that $\phi(\bar{x},y^o) > \phi(x^o,y^o) + <\theta(x^o),\bar{x}-x^o> =$ $\phi(x^o,y^o)$. This further implies that $\phi(\bar{x},y^o) > \phi(x^o,y^o) = \phi(\bar{x},\bar{y}) =$

$f_o(\bar{x}) + \langle\bar{y},f(\bar{x})\rangle = f_o(\bar{x})$. Therefore, we conclude that $\langle y^o,f(\bar{x})\rangle > 0$, but this is a contradiction, since $f(\bar{x}) \le 0$ and $y^o \ge 0$. Therefore, $x^o = \bar{x}$ and the proof is complete.

It is worthwhile mentioning that Rissanen [41] gave results which are similar to Theorems 16 and 17 above, when the convexity assumption is relaxed, but the differentiability assumption is kept. We now give the following theorem which presents a sufficient condition for the dual optimal solution to be unbounded.

Theorem 18

Let $(\bar{x},\bar{y})\epsilon Y$ such that $\theta^o(\bar{x}) + \sum_{i=1}^{m} \bar{y}_i\theta^i(\bar{x}) = 0$, where $\theta^i(\bar{x})$ is a subgradient of f_i at $\bar{x}$. If $f_i(\bar{x}) + \langle\theta^i(\bar{x}),x\rangle \le 0$ for all $i\epsilon\{1,2,\ldots,m\}$ has no solution $x\epsilon E^n$, then the dual objective function is unbounded.

Proof. Since $f_i(\bar{x}) + \langle\theta^i(\bar{x}),x\rangle \le 0$ for all $i\epsilon\{1,2,\ldots,m\}$ has no solution, then it can be shown that there exists $y \ge 0$ such that $\sum_{i=1}^{m} y_i\theta^i(\bar{x}) = 0$, and $\langle y,f(\bar{x})\rangle = 1$. See, for example, [35]. Consider $(\bar{x},\delta y+\bar{y}) \epsilon E^{n+m}$, where δ is a positive scalar,

$$\theta^o(\bar{x}) + \sum_{i=1}^{m}(\delta y_i+\bar{y}_i)\theta^i(\bar{x}) = \theta^o(\bar{x}) + \sum_{i=1}^{m}\bar{y}_i\theta^i(\bar{x}) + \delta\sum_{i=1}^{m} y_i\theta^i(\bar{x}) = 0.$$

This implies that $(\bar{x},\delta y+\bar{y})\epsilon Y$, and moreover,

$$\phi(\bar{x},\delta y+\bar{y}) = f_o(\bar{x}) + \langle\delta y+\bar{y},f(\bar{x})\rangle = f_o(\bar{x}) + \delta\langle y,f(\bar{x})\rangle + \langle\bar{y},f(\bar{x})\rangle$$

$$= \phi(\bar{x},\bar{y}) + \delta$$

This shows that ϕ is unbounded, and the proof is complete.

Finally we briefly discuss symmetric duality due to Dantzig, Cottle, and Eisenberg [13]. This duality formulation subsumes Wolfe's formulation, and consequently subsumes duality in linear and quadratic programming. They considered the following two problems.

$$\text{P: minimize}_{(x,y)}\{\psi(x,y) - \langle y,\nabla_y\psi(x,y)\rangle : x\geq 0,\ y\geq 0,\ \nabla_y\psi(x,y) \leq 0\}$$

$$\text{D: maximize}_{(x,y)}\{\psi(x,y) - \langle x,\nabla_x\psi(x,y)\rangle : x\geq 0,\ y\geq 0,\ \nabla_x\psi(x,y) \geq 0\}$$

Under the following conditions they proved certain duality results.

(i) $\psi: U\times V\to E^1$, where $E^n_+\subset U$ and $E^m_+\subset V$, and U and V are open sets.

(ii) ψ is twice continuously differentiable.

(iii) ψ is strictly convex-strictly concave on $E^n_+ \times E^m_+$.

Generalization of the above formulation to nondifferentiable functions is given by Stoer [49], Whinston [52], and Mehndiratta [38]. We now show that the symmetric duality formulation can be derived as a special case of the Minmax formulation.

Let $\phi(u,v) = \psi(x,y) - \langle x^*,x\rangle + \langle y^*,y\rangle$ where $u = (x,y^*)\epsilon E\subset E^{n+m}$, and $v = (x^*,y)\epsilon F\subset E^{n+m}$. Furthermore, let $E = F = E^n_+ \times E^m_+$. We therefore consider the functions α and β as follows.

$$\alpha(u) = \alpha(x,y^*) = \sup_{(y,x^*)}\{\psi(x,y) - \langle x^*,x\rangle + \langle y^*,y\rangle : y\geq 0,\ x^*\geq 0\}$$

$$= \sup_y\{\psi(x,y) + \langle y^*,y\rangle : y\geq 0\} \quad \text{if } x\geq 0$$

$$\infty \quad \text{if } x_i<0 \text{ for some } i\epsilon\{1,2,\ldots,m\}$$

Similarly,

$$\beta(v) = \beta(x^*,y) = \inf_{(x,y^*)}\{\psi(x,y) - \langle x^*,x\rangle + \langle y^*,y\rangle : x\geq 0,\ y^*\geq 0\}$$

$$= \inf_x\{\psi(x,y) - \langle x^*,x\rangle : x\geq 0\} \quad \text{if } y\geq 0$$

$$-\infty \qquad \text{if } y_i<0 \text{ for some } i\in\{1,2,\ldots,m\}$$

Therefore, the following two problems are subdual programs.

$$P' : \underset{x\geq 0,y^*\geq 0}{\text{minimize}}\ \sup_y\{\psi(x,y) + \langle y^*,y\rangle : y\geq 0\}$$

$$D' : \underset{y\geq 0,x^*\geq 0}{\text{maximize}}\ \inf_x\{\psi(x,y) - \langle x^*,x\rangle : x\geq 0\}$$

It may be noted that the above two problems are precisely Whinston problems for the nondifferentiable case. Under some convexity assumptions (see Theorem 1) the two problems above are dual programs. Furthermore, Stoer [49] showed that if ψ is convex-concave and differentiable on $E^n_+ \times E^m_+$, problems P' and D' are reduced to problems P and D presented earlier. It may also be noted that Balas [3] has considered a similar formulation and developed some duality results for discrete programs.

The special case when $\phi(u,v) = f_o(x) - \langle x^*,x\rangle$ has an appealing geometric interpretation. Problems P' and D' become,

$$\text{minimize}_x\{f_o(x) : x\geq 0\}, \quad \text{and}$$

$$\text{maximize}_{x^*}\{\inf_x\{f_o(x) - \langle x^*,x\rangle : x\geq 0\} : x^*\geq 0\}.$$

It may be noted that problem D' is to maximize the intercept of supporting hyperplanes of the epigraph of f_o at points corresponding to nonnegative subgradients. In the process of developing symmetric duality relationships, Whinston [52] has considered the above problem.

In closing this section it may be stated that results concerning the equivalence of optimal solutions of problems P' and D' above, and results concerning existence of solutions to these two problems can be obtained by applying Theorems 1 through 4.

CHAPTER IV

CONCLUSIONS AND RECOMMENDATIONS

Throughout this study the differentiability assumption of the functions involved in a nonlinear programming problem is relaxed. Different optimality criteria are developed, and duality in nonlinear programming is investigated. The following are the main results of this study.

1. The Fritz John necessary optimality conditions can be extended when the functions are continuous but not necessarily differentiable and locally supportable either from above or below at the point under investigation.

2. The Kuhn-Tucker necessary and sufficient conditions for optimality can be extended to the case where the functions are continuous but not necessarily differentiable and locally supportable from below at the point under consideration. The existence of an interior point of the feasible region is assumed in this case.

3. A necessary and sufficient condition for optimality is developed where neither differentiability nor continuity of the functions are assumed.

As a corollary to this result, a necessary and sufficient condition for optimality is obtained when the constraint functions are continuous and the point under investigation is an interior feasible solution.

4. Sufficient conditions for optimality are developed when the objective function is locally supportable from below and the constraint functions are not restricted. This is done by considering a new problem which is equivalent to the original problem.

5. All duality formulations in nonlinear programming can be obtained from the Minmax formulation. This unifies the apparently unrelated and diverse duality formulations.

6. By applying existing theorems of Minmax theory, different extensions of existing duality theorems are obtained.

During the course of this study several new interesting problems were encountered. The following is a brief outline of recommendations for further research in the area of optimality criteria and duality in nonlinear programming.

1. To find necessary and sufficient optimality conditions for the problem P: $\text{minimize}_x\{f_o(x)\colon g_i(x) = 0 \quad (i=1,2,\ldots,k),\ x\epsilon\Omega\}$, where $f_o,g_1,g_2,\ldots,g_k\colon E^n\to E^1$ and Ω is a nonempty set in E^n. If Ω is specialized to be the set $\{x\colon x\epsilon E^n,\ f_i(x)\le 0,\ i=1,2,\ldots,m\}$ where $f_1,f_2,\ldots,f_m\colon E^n\to E^1$, then we obtain an equality and inequality constrained program. The functions under consideration are nondifferentiable but may or may not be continuous. It may be noticed that discrete programs can be obtained by suitably specifying Ω and the functions involved in problem P above.

2. To develop solution procedures which make use of the optimality criteria developed in this study. One may construct algorithms which are validated by theorems which are in turn based on the

optimality conditions of this study (see Appendix A for example). One may also develop solution procedures which are directly based on these conditions. Finally, one may use the optimality criteria as a stopping rule or as a subroutine. It should be noted, however, that in such procedures there is an obvious drawback; namely, the process of finding a suitable subgradient of a function at a given point is not an easy task.

3. Investigating duality formulations for discrete functions in relation to the Minmax formulation.

4. Developing stronger theorems concerning existence of solutions to the two dual programs via the Minmax formulation. This can then be applied to obtain stronger results concerning different duality formulations.

5. Developing a systematic way of choosing the function ϕ and the sets E and F (see Chapter III) to obtain new duality formulations and new duality results.

APPENDIX A

A SOLUTION PROCEDURE FOR CONVEX PROGRAMS

We present a result which makes use of the generalized Kuhn-Tucker conditions developed in Chapter II, and then point out that a solution procedure for solving convex programs may be based on the result developed. We assume the problem under consideration is to minimize$_x\{f_o(x)\colon f_i(x)\leq 0,\ i=1,2,\ldots,m\}$ where, $f_o\colon E^n\to E^1$ is strictly convex function and $f_i\colon E^n\to E^1$, $i=1,2,\ldots,m$ are convex functions. It is well known that the solution to this problem, if it exists, is unique.

We now assume that $\bar{x}$ is the solution of the above problem, and consider the set of binding constraints S, i.e. $S = \{i\colon f_i(\bar{x}) = 0\}$. The following theorem asserts that the optimal solution of the equality constraint problem, minimize$_x\{f_o(x)\colon f_i(x) = 0,\ i\in S\}$ is unique, namely $\bar{x}$.

<u>Theorem 1</u>

Let $\bar{x}$ solve the problem: minimize$_x\{f_o(x)\colon f_i(x)\leq 0,\ i=1,2,\ldots,m\}$ where $f_o\colon E^n\to E^1$ is strictly convex and $f_1,\ldots,$ and $f_m\colon E^n\to E^1$ are convex. Let $S = \{i\colon f_i(\bar{x}) = 0\}$, and suppose that x^S solves the problem: minimize$_x\{f_o(x)\colon f_i(x)=0,\ i\in S\}$. Then $x^S = \bar{x}$.

Proof. Note that $\bar{x}$ is a feasible solution of the problem minimize$_x\{f_o(x)\colon f_i(x)=0,\ i\in S\}$ and hence $f_o(\bar{x}) \geq f_o(x^S)$. If we show that x^S is a feasible solution of the problem, minimize$_x\{f_o(x)\colon f_i(x)\leq 0,\ i=1,2,\ldots,m\}$, then we conclude that $f_o(x^S) \geq f_o(\bar{x})$ and hence

$f_o(\bar{x}) = f_o(x^S)$. But since the solution of the inequality constrained problem is unique then $x^S = \bar{x}$. Assume on the contrary that x^S is not feasible, then we assert that there exists $\bar{\lambda}>0$ such that $x_\lambda = (1-\lambda)\bar{x} + \lambda x^S$ is feasible for all $\lambda\epsilon[0,\bar{\lambda}]$ and not feasible for $\lambda>\bar{\lambda}$. We show this by construction. Let $\{k_1,k_2,\ldots,k_r\} \neq \phi$† be the set of indices such $f_{k_i}(\bar{x}) < 0$, $i=1,2,\ldots,r$. Let $g_{k_i}(\lambda) = f_{k_i}((1-\lambda)\bar{x} + \lambda x^S)$, for $i=1,2,\ldots,r$ and notice that,

(i) g_{k_i} is continuous for $\lambda\epsilon[0,1]$, $i=1,2,\ldots,r$.

(ii) $g_{k_i}(0) = f_{k_i}(\bar{x}) < 0$, $i=1,2,\ldots,r$.

We consider the following two cases. Case 1 corresponds to $g_{k_i}(1) = f_{k_i}(x^S) \leq 0$, which further implies that $g_{k_i}(\lambda) \leq 0$ for all $\lambda\epsilon[0,1]$. We let $\lambda_{k_i} = 1$. In Case 2, $g_{k_i}(1) = f_{k_i}(x^S) > 0$. By continuity of g_{k_i}, and since $g_{k_i}(0) < 0$, then for some $\lambda_{k_i}\epsilon(0,1)$, $g_{k_i}(\lambda_{k_i}) = 0$. We let $\bar{\lambda} = \min_i\{\lambda_{k_i} : i\epsilon\{1,2,\ldots,r\}\}$. Notice that $\bar{\lambda}\epsilon(0,1)$. It follows that $f_{k_i}(x_{\bar{\lambda}}) = g_{k_i}(\bar{\lambda})\leq 0$ for all $i\epsilon\{1,2,\ldots,r\}$ where $x_{\bar{\lambda}} = (1-\bar{\lambda})\bar{x} + \bar{\lambda}x^S$. By convexity of f_i, and since $f_j(x^S) = f_j(\bar{x}) = 0$, then $f_j(x_{\bar{\lambda}}) \leq 0$ for each $j\notin\{k_1,\ldots,k_r\}$. This implies that $x_{\bar{\lambda}}$ is feasible. By strict convexity of f_o, $f_o(x_{\bar{\lambda}}) = f_o((1-\bar{\lambda})\bar{x}+\bar{\lambda}x^S)<(1-\bar{\lambda})f_o(\bar{x})+\bar{\lambda}f_o(x^S)\leq f_o(\bar{x})$, a contradiction. Hence x^S is feasible and the proof is complete.

The following theorem gives a necessary condition that $\bar{x}$ solves the inequality constrained problem above.

---

†If $\{k_1,k_2,\ldots,k_r\} = \phi$, then the assertion follows trivially.

Theorem 2†

Let $\bar{x}$ solve the problem: minimize$_x\{f_o(x)\colon f_i(x)\le 0,\ i=1,2,\ldots,m\}$ where $f_o\colon E^n\to E^1$ is strictly convex and $f_1,f_2,\ldots,$ and $f_m\colon E^n\to E^1$ are convex. Suppose that there exists an $x\in E^n$ such that $f_i(x) < 0$ for all $i\in\{1,2,\ldots,m\}$. Let $\bar{S} = \{i\colon f_i(\bar{x}) = 0\} \neq \phi$ and let $S\subset\bar{S}$, $S\neq\bar{S}$. Then $h\in V(x^S)$ for some $h\in\bar{S} - S$, where x^S solves the problem: minimize$_x\{f_o(x)\colon f_i(x)=0,\ i\in S\}$, and $V(x^S) = \{i\colon f_i(x^S)>0\}$.

Proof. The hypotheses of Theorem 14 in Chapter II are satisfied and hence $\theta^o(\bar{x}) + \sum_{i\in\bar{S}} y_i\theta^i(\bar{x}) = 0$ where $y_i\ge 0$ for all $i\in\bar{S}$ and $\theta^i(\bar{x})$ is a subgradient of f_i at $\bar{x}$. By strict convexity of f_o it follows that,

$$f_o(x^S) > f_o(\bar{x}) + \langle\theta^o(\bar{x}),x^S-\bar{x}\rangle.$$

But since $f_i(\bar{x}) = 0$ for all $i\in S$, then $f_o(\bar{x}) \ge f_o(x^S)$ and hence it follows that $\langle\theta^o(\bar{x}),x^S-\bar{x}\rangle < 0$, which again implies that,

$$\langle \sum_{i\in S} y_i\theta^i(\bar{x}),x^S-\bar{x}\rangle + \langle \sum_{i\in\bar{S}-S} y_i\theta^i(\bar{x}),x^S-\bar{x}\rangle > 0$$

But if $i\in S$, then $f_i(\bar{x}) = f_i(x^S) = 0$, and by convexity of f_i, $f_i(x^S) \ge f_i(\bar{x}) + \langle\theta^i(\bar{x}),x^S-\bar{x}\rangle$. Moreover, since $y_i \ge 0$, we conclude that $\langle \sum_{i\in S} y_i\theta^i(\bar{x}),x^S-\bar{x}\rangle \le 0$ and hence it follows that $\sum_{i\in\bar{S}-S} y_i\langle\theta^i(\bar{x}),x^S-\bar{x}\rangle > 0$. But since $y_i\ge 0$, then we conclude that for

---

†Theorem 2 is true when f_o is strictly supportable from below at $\bar{x}$ and $f_1,f_2,\ldots,$ and f_m are supportable from below at $\bar{x}$. The proof is identical to that given.

some $h \epsilon \bar{S}-S$, $<\theta^h(\bar{x}), x^S-\bar{x}> > 0$, $y_h>0$. By convexity of f_h, then

$$f_h(x^S) \geq f_h(\bar{x}) + <\theta^h(\bar{x}), x^S-\bar{x}>.$$

But since $f_h(\bar{x}) = 0$, then $f_h(x^S) > 0$ and the proof is complete.

Corollary

Let $\bar{x}$ solve the above inequality problem. Then $f_h(x^{\bar{S}-\{h\}}) > 0$ for all $h \epsilon \bar{S}$.

This corollary is an immediate consequence of the theorem where we let $S = \bar{S} - \{h\}$.

The above theorem suggests a solution procedure for solving convex programs, by solving a sequence of equality constrained problems. A "Branch and Bound" algorithm may be adopted, where branching is done according to the corollary to Theorem 2, and bounding is done by solving equality constrained problems. For example, given a solution x^S of the problem, minimize$_x\{f_o(x)$: $f_i(x) = 0$, $i \epsilon S\}$ which violates the constraints $k_1, k_2, \ldots, k_r$, then branching from x^S leads to r problems, to minimize$_x\{f_o(x)$: $f_i(x) = 0$, $i \epsilon S \cup \{k_j\}\}$ for $j=1,2,\ldots,r$. The optimality criteria due to Branch and Bound [2], or the generalized Kuhn-Tucker conditions can be adopted as a stopping rule. The efficiency of the procedure is doubtful, however, due to the combinatorial nature of the method and the need for solving a sequence of equality constrained problems.

APPENDIX B

THE EXAMPLE OF KUHN AND TUCKER

We now consider two examples. The first example is the well known example of the outward cusp due to Kuhn and Tucker [30]. This example does not satisfy the constraint qualification, but we show that it satisfies our sufficient condition and hence optimality is assured. We give another example where due to lack of convexity no claim of optimality can be made. However, the constraint qualification is satisfied at the point under consideration. We show that our sufficient condition is satisfied and hence optimality is assured.

The Kuhn-Tucker example is to minimize$_x\{f_o(x_1,x_2)$: $f_1(x_1,x_2) \leq 0$, $f_2(x_1,x_2) \leq 0$, $f_3(x_1,x_2) \leq 0\}$ where,

$$f_o(x_1,x_2) = -x_1$$

$$f_1(x_1,x_2) = x_2 - (1-x_1)^3$$

$$f_2(x_1,x_2) = -x_2$$

$$f_3(x_1,x_2) = -x_1$$

Consider Figure 6 which shows the feasible region S. It is obvious that the optimal solution is -1 and is attained at $\bar{x} = (1,0)$.

The binding constraints at $\bar{x}$ are constraint 1 and constraint 2. We show that the Kuhn-Tucker constraint qualification is not satisfied. We calculate $\nabla f_1(\bar{x})$ and $\nabla f_2(\bar{x})$ as follows,

$$\nabla f_1(x_1,x_2) = (3(1-x_1)^2,1), \text{ and hence } \nabla f_1(\bar{x}) = (0,1).$$

$$\nabla f_2(x_1,x_2) = (0,-1).$$

It is immediate that the vector $h = (1,0)$ satisfies the two conditions $\langle h,\nabla f_1(\bar{x})\rangle = \langle h,\nabla f_2(\bar{x})\rangle = 0$. However, there exists no differentiable arc $a(\theta)$ which is contained in the feasible region such that $\left[\frac{da(\theta)}{d\theta}\right]_{\theta=0} = \alpha h$ where α is a positive scalar, and $a(0) = \bar{x}$.

We now show that one of our sufficient conditions is satisfied at $\bar{x} = (1,0)$. Consider the set $A_o = \{(x_1,x_2,y)\colon y \geq f_o(x_1,x_2) = -x_1\}$. We show that $(-1,0,-1)$ is an *o.n.* to A_o at $(\bar{x}_1,\bar{x}_2,f_o(\bar{x})) = (1,0,-1)$. Let $(x_1,x_2,y) \in A_o$, then $\langle(-1,0,-1), (x_1,x_2,y) - (1,0,-1)\rangle = -x_1-y\leq 0$. Hence $(\theta^o(\bar{x}),-1) = (-1,0,-1)$ and therefore $\theta^o(\bar{x}) = (-1,0)$. But $-\theta^o(\bar{x}) = (1,0)$ is an *o.n.* to the feasible region S at $\bar{x}$, and hence by Theorem 17 of Chapter II, it follows that $\bar{x}$ is an optimal solution of the problem.

Consider another example which satisfies the Kuhn-Tucker constraint qualification. Consider the problem, minimize$_x\{f_o(x_1,x_2)\colon$ $f_1(x_1,x_2)\leq 0$, $f_2(x_1,x_2)\leq 0$, $f_3(x_1,x_2)\leq 0$, $f_4(x_1,x_2)\leq 0\}$, where,

$$f_o(x_1,x_2) = -x_1$$

$$f_1(x_1,x_2) = 1 - (x_1-2)^2 + x_2$$

$$f_2(x_1,x_2) = x_1^2 + x_2^2 - 1$$

$$f_3(x_1,x_2) = -x_2$$

$$f_4(x_1,x_2) = -x_1$$

From Figure 7 it is clear that the optimal solution is -1 and is attained at $\bar{x} = (\bar{x}_1,\bar{x}_2) = (1,0)$. The Kuhn-Tucker constraint qualification is satisfied at $\bar{x}$. We show that the Kuhn-Tucker conditions, $\nabla f_o(\bar{x}) + \sum_{i=1}^{3} u_i \nabla f_i(\bar{x}) = 0$ are satisfied. The gradients ∇f_o, ∇f_1, ∇f_2, and ∇f_3 are calculated as follows,

$$\nabla f_o(x_1,x_2) = (-1,0)$$

$$\nabla f_1(x_1,x_2) = (-2(x_1-2),1), \text{ and hence } \nabla f_1(\bar{x}) = (2,1)$$

$$\nabla f_2(x_1,x_2) = (2x_1,2x_2), \quad \text{ and hence } \nabla f_2(\bar{x}) = (2,0)$$

$$\nabla f_3(x_1,x_2) = (0,-1)$$

Therefore, the Kuhn-Tucker conditions are satisfied by letting $u_1 = 1/2$.

$u_2 = 0$, $u_3 = 1/2$. However, we cannot claim that $\bar{x}$ is an optimal point since f_1 is not a convex function.

However, we show that $\bar{x}$ is an optimal point since it satisfies our sufficient conditions. As before, $(-1,0,-1)$ is an *o.n.* to A_o at $(\bar{x},f_o(\bar{x}))$, and since $-\theta^o(\bar{x}) = (1,0)$ is an *o.n.* to S at $\bar{x}$, then $\bar{x}$ is an optimal solution.

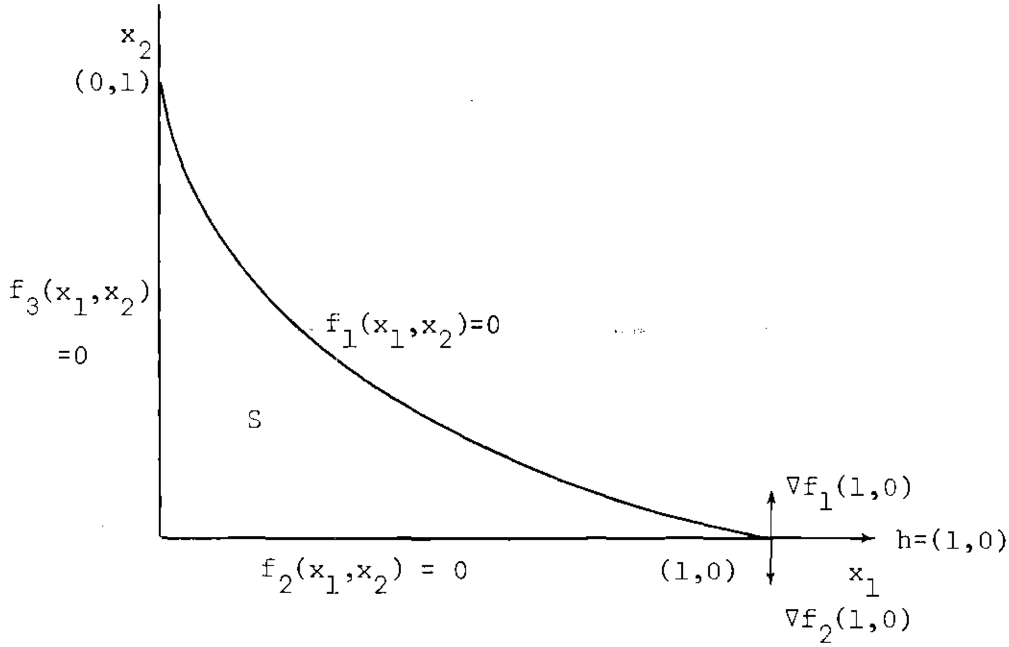

Figure 6. The Example of Kuhn and Tucker

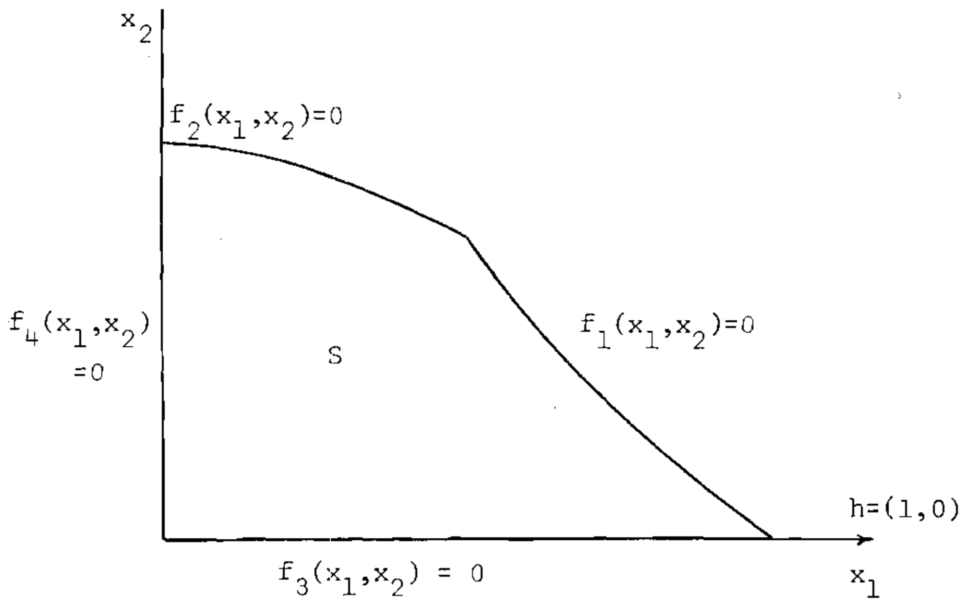

Figure 7. An Application of the Sufficiency Conditions

BIBLIOGRAPHY

1. Arrow, K. J., L. Hurwicz, and H. Uzawa, "Constraint Qualifications in Maximization Problems," *Naval Research Logistics Quarterly*, 8, 1961.

2. Balas, E., "A Note on the Branch and Bound Principle," *Operations Research*, 16, 1968.

3. Balas, E., *Minimax and Duality for Linear and Nonlinear Mixed-Integer Programming*, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1969.

4. Barankin, E., and R. Dorfman, "A Method for Quadratic Programming," *Econometrica*, 24, 1956.

5. Berge, C., and A. Ghoulia-Houri, *Programming, Games, and Transportation Networks*, John Wiley and Sons, Inc., New York, 1965.

6. Bracken, J., and G. P. McCormick, *Selected Applications of Nonlinear Programming*, John Wiley and Sons,Inc., New York, 1968.

7. Bram, J., "The Lagrange Multiplier Theorem for Max-Min with Several Constraints," *J. SIAM Appl. Math.*, 14, 1966.

8. Brønsted, A., and R. T. Rockafeller, "On the Subdifferentiability of Convex Functions," *Proc. Amer. Math. Soc.*, 16, 1965.

9. Canon, M., C. Cullum, and E. Polak, "Constrained Minimization Problems in Finite Dimensional Spaces," *J. SIAM Control*, 4, 1966.

10. Charnes, A., and W. W. Cooper, *Management Models and Industrial Applications of Linear Programming,"* Vol. I, John Wiley and Sons, Inc., New York, 1967.

11. Cottle, R. W., "Symmetric Dual Quadratic Programs," *Quart. Appl. Math.*, 21, 1963.

12. Danskin, J. M., *The Theory of Max-Min and its Applicants to Weapons Allocation Problems*, Springer-Verlag, New York, 1967.

13. Dantzig, G. B., E. Eisenberg, and R. W. Cottle, "Symmetric Dual Nonlinear Programs," *Pacific Journal of Mathematics*, 15, 1965.

14. Dorn, W. S., "A Duality Theorem for Convex Programming," *IBM J. Res. Develop.*, 4, 1960.

15. Egglestone, H. G., *Convexity,* Cambridge University Press, 1958.

16. Everett, H., "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," *Operations Research,* 11, 1963.

17. Falk, J. E., "Lagrange Multipliers and Nonlinear Programming," *J. Math. Anal. Appl.,* 19, 1967.

18. Fenchel, W., "On Conjugate Convex Functions," *Canad. J. Math.*, 1, 1949.

19. Fenchel, W., "Convex Cones, Sets, and Functions," Lecture Notes (Mimeographed), Princeton University, 1953.

20. Fiacco, A. V., and G. P. McCormick, "The Sequential Unconstrained Minimization Technique for Nonlinear Programming: A Primal-Dual Method," *Management Science,* 10, 1964.

21. Fiacco, A. V., and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques,* John Wiley and Sons, Inc., New York, 1968.

22. Gale, D., *The Theory of Linear Economic Models,* McGraw-Hill Book Company, Inc., 1960.

23. Hadley, G., *Nonlinear and Dynamic Programming,* Addison-Wesley, Reading, Mass., 1964.

24. Hildreth, C. G., "A Quadratic Programming Procedure," *Naval Research Logistics Quarterly,* 4, 1957.

25. Hirsch, W. M., and G. B. Dantzig, "The Fixed Charge Problem," *The Rand Corporation,* Paper P-648, 1954.

26. Huard, P., "Dual Programs," *IBM J. Res. Develop.,* 6, 1962.

27. John, F., "Extremum Problems with Inequalities as Side Conditions," in K. O. Friedrichs, O. E. Neugebauer, and J. J. Stoker (eds.), *Studies and Essays, Courant Anniversary Volume,* Wiley (Interscience), New York, 1948.

28. Karamadian, S., "Strictly Quasi-Convex (Concave) Functions and Duality in Mathematical Programming," *J. Math. Anal. Appl.,* 20, 1967.

29. Karlin, S., *Mathematical Methods and Theory in Games, Programming, and Economics,* Vol. II, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1959.

30. Kuhn, H. W., and A. W. Tucker, "Nonlinear Programming," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951.

31. Künzi, H. P., W. Krelle, and W. Oettli, *Nonlinear Programming*, Blaisdell Publishing Company, 1966.

32. Luenberger, D. G., *Optimization by Vector Space Methods*, John Wiley and Sons, Inc., New York, 1969.

33. Mangasarian, O. L., "Duality in Nonlinear Programming," *Quart. Appl. Math.*, 20, 1962.

34. Mangasarian, O. L., "Pseudo-Convex Functions," *J. SIAM Control*, 3, 1965.

35. Mangasarian, O. L., *Nonlinear Programming*, McGraw-Hill Book Company, 1969.

36. Mangasarian, O. L., and J. Ponstein, "Minmax and Duality in Nonlinear Programming," *J. Math. Anal. Appl.*, 11, 1965.

37. Mangasarian, O. L., and S. Fromovitz, "The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints, *J. Math. Anal. Appl.*, 17, 1967.

38. Mehndiratta, S. L., "Self-Duality in Mathematical Programming," *J. SIAM Appl. Math.*, 15, 1967.

39. Mond, B., "A Symmetric Dual Theorem for Nonlinear Programs," *Quart. Appl. Math.*, 23, 1965.

40. Moreau, J. J., "Etudé Locale d'une Functionelle Convexe," *Faculté des Sciences de Montpellier*, Séminaires de Mathématiques, 1963.

41. Rissanen, J., "On Duality Without Convexity," *J. Math. Anal. Appl.*, 18, 1967.

42. Rockafeller, R. T., "Minimax Theorems and Conjugate Saddle Functions," *Math. Scand.*, 14, 1964.

43. Rockafeller, R. T., "Extension of Fenchel's Duality Theorem for Convex Functions," *Duke Math. Journal*, 33, 1966.

44. Rockafeller, R. T., "Duality in Nonlinear Programming," in G. B. Dantzig, and A. Veinott (eds.), *Mathematics of the Decision Sciences*, American Mathematical Society, 1969.

45. Roode, J. D., *Generalized Lagrangian Functions in Mathematical Programming,* Published Ph.D. Dissertation, State University of Leyden, Holland, 1968.

46. Rosen, J. B., "The Gradient Projection Method for Nonlinear Programming, Part I: Linear Constraints," *J. SIAM Appl. Math.,* 8, 1960.

47. Sion, M., "On General Minmax Theorems," *Pacific J. Math.,* 8, 1958.

48. Stoer, J., "Duality in Nonlinear Programming and the Minimax Theorem," *Numer. Math.,* 5, 1963.

49. Stoer, J., "Über einen Dualitätssatz der Nichtlinearen Programmierung," *Numer. Math.,* 6, 1964.

50. Theil, H., and C. Van de Panne, "Quadratic Programming as an Extension of Conventional Quadratic Maximization," *Management Science,* 7, 1961.

51. Valentine, F. A., *Convex Sets,* McGraw-Hill Book Company, New York, 1964.

52. Whinston, A., "Some Applications of the Conjugate Function Theory to Duality," in J. Abadie (ed.), *Nonlinear Programming,* North Holland Publishing Company, Amsterdam, 1967.

53. Wilde, D. J., and C. S. Beightler, *Foundations of Optimization,* Prentice-Hall, Inc., N. J., 1967.

54. Williams, A. C., "Nonlinear Activity Analysis," *Mich. Univ. Engineering Summer Conferences,* Recent Mathematical Advances, 1968.

55. Wolfe, P., "The Simplex Method for Quadratic Programming," *Econometrica,* 27, 1959.

56. Wolfe, P., "A Duality Theorem for Nonlinear Programming," *Quart. Appl. Math.,* 19, 1961.

57. Zangwill, W. I., "The Piecewise Concave Function," *Management Science,* 13, 1967.

58. Zangwill, W. I., *Nonlinear Programming: A Unified Approach,* Prentice-Hall, Inc., Englewood Cliffs, N. J., 1969.

59. Zoutendijk, G., *Methods of Feasible Directions,* Elsevier, Amsterdam, 1960.

VITA

The author was born in Cairo, Egypt on March 2, 1943. He attended Cairo University during the period September, 1959, to July, 1964. He had his Bachelor degree in Mechanical Engineering with the first degree of honor. The author then worked for two years as an instructor in Cairo University.

He joined Georgia Institute of Technology in September, 1966, and obtained his Master degree in Industrial Engineering in September, 1967. In 1969 the author obtained his Ph.D. from the same Institute. The dissertation is titled, "Nonlinear Programming: Nondifferentiable Functions." The author is currently working as an Assistant Professor in the School of Industrial and Systems Engineering, Georgia Institute of Technology.